(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 March 2004 (11.03.2004)

PCT

(10) International Publication Number
WO 2004/020595 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2003/027107

(22) International Filing Date: 28 August 2003 (28.08.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/406,616	29 August 2002 (29.08.2002)	US
60/406,655	29 August 2002 (29.08.2002)	US
60/406,640	29 August 2002 (29.08.2002)	US
60/406,576	29 August 2002 (29.08.2002)	US
60/406,666	29 August 2002 (29.08.2002)	US
60/406,611	29 August 2002 (29.08.2002)	US
60/406,612	29 August 2002 (29.08.2002)	US
60/411,019	17 September 2002 (17.09.2002)	US
60/411,024	17 September 2002 (17.09.2002)	US
60/411,046	17 September 2002 (17.09.2002)	US
60/411,082	17 September 2002 (17.09.2002)	US
60/411,022	17 September 2002 (17.09.2002)	US
60/410,962	17 September 2002 (17.09.2002)	US
60/410,953	17 September 2002 (17.09.2002)	US
60/410,957	17 September 2002 (17.09.2002)	US
60/411,037	17 September 2002 (17.09.2002)	US
60/410,951	17 September 2002 (17.09.2002)	US
60/410,946	17 September 2002 (17.09.2002)	US
60/410,960	17 September 2002 (17.09.2002)	US
60/411,111	17 September 2002 (17.09.2002)	US
60/411,052	17 September 2002 (17.09.2002)	US

(71) Applicants *(for all designated States except US)*: **FIVE PRIME THERAPEUTICS, INC.** [US/US]; 951 Gateway Boulevard, South San Francisco, CA 94080 (US). **RIKEN The Institute of Physical and Chemical Research** [JP/JP]; 2-1, Hirosawa, Wako,Saitama, 351-0198 (JP). **KABUSHIKI KAISHA DNAFORM** [JP/JP]; 3-35, Mita 1-chome, Minato-ku, Tokyo 108-0073 (JP).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WILLIAMS, Lewis, T.** [US/US]; 125 Chapel Drive, Mill Valley, CA 94941 (US). **CHU, Keting** [US/US]; 2017 Easton Drive, Burlingame, CA 94010 (US). **LEE, Ernestine** [US/US]; 226 Arlington Avenue, kensington, CA 94707 (US). **HESTIR, Kevin** [/]; 226 Arlington Avenue, Kensington, CA 94707 (US).

(74) Agent: **GARRETT, Arthur, S.**; Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P., 1300 I Street, NW, Washington, DC 20005-3315 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL HUMAN POLYPEPTIDES ENCODED BY POLYNUCLEOTIDES

(57) Abstract: The invention provides novel polynucleotides, related polypeptides, related nucleic acid and polypeptide compositions, and related modulators, such as antibodies and small molecule modulators. The compositions of the invention are useful in treating proliferative disorders, e.g., cancers, and inflammatory, immune, bacterial, and viral disorders.

WO 2004/020595 A2

NOVEL HUMAN POLYPEPTIDES ENCODED BY POLYNUCLEOTIDES

PRIORITY CLAIM

[001] This application is related to the following provisional applications filed in the United States Patent and Trademark Office, the disclosures of which are hereby incorporated by reference:

Application Number	Title	Filing Date
60/406,616	Polynucleotides Encoding Secreted Proteins and Secreted Proteins Encoded Thereby	August 29, 2002
60/406,655	Polynucleotides Encoding Single Transmembrane Proteins And Single Transmembrane Proteins Encoded Thereby	August 29, 2002
60/406,640	Polynucleotides Encoding Multiple Transmembrane Proteins And Multiple Transmembrane Proteins Encoded Thereby	August 29, 2002
60/406,576	Polynucleotides Encoding Kinases and Kinases Encoded Thereby	August 29, 2002
60/406,666	Polynucleotides Encoding Proteases and Proteases Encoded Thereby	August 29, 2002
60/406,611	Polynucleotides Encoding Phosphatases and Phosphatases Encoded Thereby	August 29, 2002
60/406,612	Polynucleotides Encoding Polypeptides and Polypeptides Encoded Thereby	August 29, 2002
60/411,019	Polynucleotides Encoding Secreted Proteins and Secreted Proteins Encoded Thereby	September 17, 2002
60/411,024	Novel Polynucleotides Encoding Secreted Proteins and Novel Secreted Proteins Encoded Thereby	September 17, 2002
60/411,046	Polynucleotides Encoding Single Transmembrane Proteins and Single Transmembrane Proteins Encoded Thereby	September 17, 2002

60/411,082	Novel Polynucleotides Encoding Single Transmembrane Proteins and Novel Single Transmembrane Proteins Encoded Thereby	September 17, 2002
60/411,022	Polynucleotides Encoding Multiple Transmembrane Proteins and Multiple Transmembrane Proteins Encoded Thereby	September 17, 2002
60/410,962	Novel Polynucleotides Encoding Multiple Transmembrane Proteins and Novel Multiple Transmembrane Proteins Encoded Thereby	September 17, 2002
60/410,953	Polynucleotides Encoding Kinases And Kinases Encoded Thereby	September 17, 2002
60/410,957	Novel Polynucleotides Encoding Kinases and Novel Kinases Encoded Thereby	September 17, 2002
60/411,037	Polynucleotides Encoding Phosphatases and Phosphatases Encoded Thereby	September 17, 2002
60/410,951	Novel Polynucleotides Encoding Phosphatases and Novel Phosphatases Encoded Thereby	September 17, 2002
60/410,946	Polynucleotides Encoding Proteases and Proteases Encoded Thereby	September 17, 2002
60/410,960	Novel Polynucleotides Encoding Proteases and Novel Proteases Encoded Thereby	September 17, 2002
60/411,111	Polynucleotides Encoding Polypeptides and Polypeptides Encoded Thereby	September 17, 2002
60/411,052	Novel Polynucleotides Encoding Proteins and Novel Proteins Encoded Thereby	September 17, 2002

TECHNICAL FIELD

[002] The present invention is related generally to novel polynucleotides and novel polypeptides encoded thereby, their compositions, antibodies directed thereto, and other agonists or antagonists thereto. The polynucleotides and polypeptides are useful in diagnostic, prophylactic, and therapeutic applications for a variety of diseases, disorders, syndromes and

conditions, as well as in discovering new diagnostics, prophylactics, and therapeutics for such diseases, disorders, syndromes, and conditions (hereinafter disorders).

[003] This application further relates to the field of polypeptides that are associated with regulating cell growth and differentiation, that are over-expressed in cancer, and/or that can be associated with proliferation or inhibition of cancer growth, including hematopoietic cancers such as leukemias, lymphomas, and solid cancers such as lung cancer, for example, adenocarcinomas and/or squamous cell carcinomas. These polypeptides may also be associated with other conditions, such as inflammatory, immune, and metabolic disorders, as well as microbial infections, including viral, bacterial, fungal, and parasitic diseases, disorders, syndromes, or conditions.

[004] This application further relates to modulators of biological activity that can specifically bind to these polynucleotides or polypeptides, or otherwise specifically modulate their activity. For example, they can directly or indirectly induce antibody-dependent cellular cytotoxicity (ADCC), complement-dependent cytotoxicity (CDC), endocytosis, apoptosis, or recruitment of other cells to effect cell activation, cell inactivation, cell growth or differentiation or inhibition thereof, and cell killing.

[005] The sequences of the invention encompass a variety of different types of nucleic acids and polypeptides with different structures and functions. They can encode or comprise polypeptides belonging to different protein families ("Pfam"). The "Pfam" system is an organization of protein sequence classification and analysis, based on conserved protein domains; it can be publicly accessed in a number of ways, for example, at http://pfam.wustl.edu. Protein domains are portions of proteins that have a tertiary structure and sometimes have enzymatic or binding activities; multiple domains can be connected by flexible polypeptide regions within a protein. Pfam domains can comprise the N-terminus or the C-terminus of a protein, or can be situated at any point in between. The Pfam system identifies protein families based on these domains and provides an annotated, searchable database that classifies proteins into families (Bateman et al., 2002).

[006] Sequences of the invention can encode or be comprised of more than one Pfam. Sequences encompassed by the invention include, but are not limited to, the polypeptide and polynucleotide sequences of the molecules shown in the Sequence Listing and corresponding molecular sequences found at all developmental

stages of an organism. Sequences of the invention can comprise genes or gene segments designated by the Sequence Listing, and their gene products, i.e., RNA and polypeptides. They also include variants of those presented in the Sequence Listing that are present in the normal physiological state, e.g., variant alleles such as SNPs, splice variants, as well as variants that are affected in pathological states, such as disease-related mutations or sequences with alterations that lead to pathology, and variants with conservative amino acid changes. Sequences of the invention are categorized below; any given sequence can belong to one or more than one category.

Secreted Protein-Related Sequences

[007] Secreted proteins, also referred to as secreted factors, include proteins that are produced by cells and exported extracellularly, extracellular fragments of transmembrane proteins that are proteolytically cleaved, and extracellular fragments of cell surface receptors, which fragments may be soluble. An example of a secreted protein is keratinocyte growth factor (KGF), which stimulates the growth of keratinocytes, and is useful for repairing tissue after chemotherapy or radiotherapy.

[008] Many and widely variant biological functions are mediated by a wide variety of different types of secreted proteins. Yet, despite the sequencing of the human genome, relatively few pharmaceutically useful secreted proteins have been identified. It would be advantageous to discover novel secreted proteins or polypeptides, and their corresponding polynucleotides that have medical utility.

[009] Pharmaceutically useful secreted proteins of the present invention will have in common the ability to act as ligands for binding to receptors on cell surfaces in ligand/receptor interactions, to trigger certain intracellular responses, such as inducing signal transduction to activate cells or inhibit cellular activity, to induce cellular growth, proliferation, or differentiation, or to induce the production of other factors that, in turn, mediate such activities.

[010] The cell types having cell surface receptors responsive to secreted proteins are various, including, for example, stem cells; progenitor cells; and precursor cells and mature cells of the hematopoietic, hepatic, neural, lung, heart, thymic, splenic, epithelial, pancreatic, adipose, gastrointestinal, colonic, optic, olfactory, bone and musculoskeletal lineages. Further, the hematopoietic cells can be red blood cells or white blood cells, including cells of the B lymphocytic (B cell), T lymphocytic (T cell), dendritic, megakaryocytic, natural killer (NK), macrophagic, eosinophilic, and basophilic lineages. The cell types responsive to secreted proteins

also include normal cells or cells implicated in disorders or other pathological conditions.

[011] As an example, certain of the secreted proteins of the present invention can stimulate T or B cell growth or differentiation by interacting with precursor T or B cells or hematopoietic progenitor cells, or bone marrow stem cells. As another example, certain secreted proteins of the present invention can maintain stem cells, progenitor cells or precursor cells in an undifferentiated state. As a further example, certain secreted proteins of the present invention can regulate bone growth by stimulation or inhibition thereof, secretion of insulin, glucose metabolism, cell proliferation, response to microbial infection, and regeneration of tissues including neural, muscular, and epithelial. Moreover, certain secreted proteins of the present invention can induce apoptosis such as in cancer cells or inflammatory cells.

[012] Certain of the secreted proteins of the present invention are useful for diagnosis, prophylaxis, or treatment of disorders, in subjects that are deficient in such secreted proteins or require regeneration of certain tissues, the proliferation of which is dependent on such secreted proteins, or requires an inhibition or activation of growth that is dependent on such secreted proteins. Examples of such disorders include cancer, such as bone cancer, brain tumors, breast and ovarian cancer, Burkitt's lymphoma, chronic myeloid leukemia, colon cancer, endocrine system cancers, gastrointestinal cancers, gynecological cancers, head and neck cancers, leukemia, lung cancer, lymphomas, malignant melanoma, metastases, multiple endocrine neoplasia, myelomas, neurofibromatosis, pancreatic cancer, pediatric cancers, penile cancer, prostate cancer, disorders related to the Ras oncogene, retinoblastoma (RB), sarcomas, skin cancers, testicular cancer, thyroid cancer, urinary tract cancers, and von Hippel-Lindau syndrome.

[013] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of hematopoeisis, including thrombosis; bleeding; anemias, e.g., iron deficiency and other hypoproliferative anemias, megaloblastic anemias, hemolytic anemias, acute blood loss, and aplastic anemia; hemoglobinopathies; disorders of granulocytes and monocytes; myelodysplasias and related bone marrow failure syndromes; polycythemias, e.g., polycythemia vera; acute and chronic myeloid leukemia, and other myeloproliferative diseases, e.g., malignancies of lymphoid cells; stimulation of replacement cell growth following irradiation or chemotherapy; and plasma cell disorders.

[014] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of hemostasis, such as disorders of the platelet and vessel wall, disorders of coagulation and thrombosis, and anticoagulant, fibrinolytic and antiplatelet therapies.

[015] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the cardiovascular system including disorders of the heart, such as heart failure; congenital heart disease; rheumatic fever; cor pulmonale; cardiomyopathies e.g., myocarditis; pericardial disease; cardiac tumors; cardiac manifestations of systemic diseases; and vascular diseases, such as acute myocardial infarction, ischemic heart disease, hypertensive vascular disease, diseases of the aorta, and vascular diseases of the extremities.

[016] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the respiratory system, such as asthma, hypersensitivity pneumonitis, e.g., with pulmonary infiltration, pneumonia, necrotizing pulmonary infections, bronchiectasis, cystic fibrosis, chronic bronchitis, emphysema and airway obstruction, interstitial lung diseases, primary pulmonary hypertension, pulmonary thromboembolism, disorders of the pleura, mediastinum, and diaphragm, disorders of ventilation, sleep apnea, and acute respiratory distress syndrome.

[017] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the kidney and urinary tract, such as, for example, chronic renal failure and glomerulopathies.

[018] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the gastrointestinal system, including disorders of the alimentary tract, such as, for example, peptic ulcer disease and related disorders, inflammatory bowel disease, irritable bowel syndrome; disorders of the liver and biliary tract, such as, for example, hyperbilirubinemias, acute viral hepatitis, chronic hepatitis, and cirrhosis; and disorders of the pancreas, such as acute or chronic pancreatitis.

[019] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the immune system, connective tissue, and joints, including, for example, autoimmune diseases, primary immune deficiency diseases, human immunodeficiency virus diseases, allergies, systemic lupus erythematosus, rheumatoid arthritis, systemic sclerosis, Sjogren's syndrome,

ankylosing spondylitis, reactive arthritis, vasculitis, sarcoidosis, amyloidosis, osteoarthritis, gout, psoriatic, and other arthritis.

[020] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the endocrine system, including, for example, disorders of the pituitary, hypothalamus, neurohypophysis, thyroid gland, adrenal cortex, testes, ovary, and other organs of the female reproductive system, such as breast; as well as pheochromocytoma, diabetes mellitus, and hypoglycemia.

[021] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of bone and mineral metabolism, and other metabolic processes, including, for example, diseases of the parathyroid gland and other hyper- and hypocalcemic disorders, osteoporosis, Paget's disease and other dysplasia of bone, disorders of lipoprotein metabolism, hemochromatosis, porphyries, disorders of purine and pyrimidine metabolism, Wilson's disease, lysosomal storage diseases, glycogen storage diseases, lipodystrophies, and other primary disorders of adipose tissue.

[022] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the central nervous system, including, for example, seizures and epilepsy, cerebrovascular diseases, Alzheimer's disease and other extrapyramidal disorders, ataxic disorders, amylotrophic lateral sclerosis and other motor neuron diseases, disorders of the autonomic nervous system, diseases of the spinal cord, including spinal cord injury, primary and metastatic tumors of the nervous system, multiple sclerosis, and other demyelinating diseases, as well as chronic and recurrent meningitis.

[023] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of nerves or muscle, including, for example, Guillain-Barre Syndrome, myasthenia gravis and other diseases of the neuromuscular junction, polymyositis, dermatomyositis, muscular dystrophies, and other muscle diseases.

[024] Certain of the secreted proteins herein can be used for diagnosis, prophylaxis, and treatment of disorders of the skin, including, for example, eczema, psoriasis, cutaneous infections, acne, and other common skin disorders, and immunologically mediated skin diseases.

[025] The agonists or antagonists of the secreted proteins herein or fragments thereof can be useful in treating elevated levels of such proteins in ny of the

disorders above, and including angina, anoxia, arrhythmias, asthma, atherosclerosis, benign prostatic hyperplasia, Buerger's Disease, cardiac arrest, cardiogenic shock, cerebral trauma, Crohn's Disease, congenital heart disease, mild congestive heart failure (CHF), severe congestive heart failure, cerebral ischemia, cerebral infarction, cerebral vasospasm, cirrhosis, diabetes, dilated cardiomyopathy, endotoxic shock, gastric mucosal damage, glaucoma, head injury, hemodialysis, hemorrhagic shock, hypertension (essential), hypertension (malignant), hypertension (pulmonary), hypertension (e.g., pulmonary, after bypass), hypoglycemia, inflammatory arthritis, ischemic bowel disease, ischemic disease, male penile erectile dysfunction, malignant hemangioendothelioma, myocardial infarction, myocardial ischemia, prenatal asphyxia, postoperative cardiac surgery, prostate cancer, preeclampsia, Raynaud's Phenomenon, renal failure (acute), renal failure (chronic), renal ischemia, restenosis, sepsis syndrome, subarachnoid hemorrhage (acute), surgical operations, status epilepticus, stroke (thromboembolic), stroke (hemorrhagic), Takayasu's arteritis, ulcerative colitis, uremia after hemodialysis, and uremia before hemodialysis.

[026] Secreted proteins can be screened for functional activities in appropriate functional assays, as is conventional in the art. Such assays include, for example, *in vitro* and *in vivo* assays for factors that stimulate the proliferation or differentiation of stem cells, progenitor cells, or precursor cells into T cells, B cells, pancreatic islet cells, bone cells, neuronal cells, etc.

[027] The tetratricopeptide repeat (TPR) is an example of a protein domain characteristic of a protein family, and is present in some of the secreted polypeptides of the invention. The TPR family is characterized by a degenerate 34 amino acid sequence present in a wide variety of proteins; it mediates protein-protein interactions, and is involved in scaffold formation and the assembly of multiprotein complexes (http://pfam.wustl.edu/cgi-bin/getdesc?name=TPR). Secreted protein-related sequences can also possess or interact with cytochrome P450 domains, which are involved in the oxidative degradation of various compounds, including environmental toxins and mutagens (http://pfam.wustl.edu/cgi-bin/getdesc?name=p450). Secreted protein-related sequences, e.g., cholesteryl ester transfer protein and phospholipid transfer protein, can also possess or interact with the LBP/BPI/CETP domain, which is characteristically found in lipid-binding serum glycoproteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=LBP_BPI_CETP). Secreted protein-related sequences can also possess or interact with peptidase S8 domains, also known as subtilase domains,

which are comprised of serine proteases with a wide range of peptidase activities, including exopeptidase, endopeptidase, oligopeptidase, and omega-peptidase activity (http://pfam.wustl.edu/cgi-bin/getdesc?name=Peptidase_S8).. Secreted protein-related sequences can also possess or interact with adh_short, or short-chain dehydrogenase domains, which are found in a large family of proteins, and are made up of short-chain dehydrogenases and reductase enzymes; most family members function as NAD- or NADP- dependent oxidoreductases (http://pfam.wustl.edu/cgi-bin/getdesc?name=adh_short).

[028] The inventors herein have identified novel secreted proteins using an algorithm that is constructed on the basis of a number of attributes including hydrophobicity, two-dimensional structure, prediction of signal sequence cleavage site, and other parameters. Based on such algorithm, a sequence that has a secreted tree vote of 0.5 – 1.0, preferably, 0.6 – 1.0, is believed to be a secreted protein.

Transmembrane Protein-Related Sequences

[029] Transmembrane proteins extend into or through the cell membrane's lipid bilayer; they can span the membrane once, or more than once. Transmembrane proteins that span the membrane once are "single transmembrane proteins" (STM), and transmembrane proteins that span the membrane more than once are "multiple transmembrane proteins" (MTM). Examples of transmembrane proteins include the insulin receptor, adenylate cyclase, and intestinal brush border esterase.

[030] A single transmembrane protein typically has one transmembrane (TM) domain, spanning a series of consecutive amino acid residues, numbered on the basis of distance from the N-terminus, with the first amino acid residue at the N-terminus as number 1. A multi-transmembrane protein typically has more than one TM domain, each spanning a series of consecutive amino acid residues, numbered in the same way as the STM protein.

[031] Transmembrane proteins, having part of their molecules on either side of the bilayers, have many and widely variant biological functions. They transport molecules, e.g., ions or proteins across membranes, transduce signals across membranes, act as receptors, and function as antigens. Transmembrane proteins are often involved in cell signaling events; they can comprise signaling molecules, or can interact with signaling molecules. For example, tyrosine kinases can be transmembrane receptor proteins. Abnormalities of receptor tyrosine kinases are associated with human cancers; tumor cells are known to use receptor tyrosine kinases

in transduction pathways to achieve tumor growth, angiogenesis and metastasis. Therefore, receptor tyrosine kinases represent pivotal targets in cancer therapy. It would be similarly advantageous to discover novel transmembrane proteins or polypeptides, and their corresponding polynucleotides that have additional medical utility.

[032] The transmembrane polypeptides of the invention, like the secreted polypeptides, also have many different functional domains, and belong to a wide variety of Pfam families. Transmembrane protein-related sequences can possess or interact with immunoglobulin (ig) domains, which are characteristically found in the immunoglobulin superfamily, comprised of hundreds of proteins, with various functions (http://pfam.wustl.edu/cgi-bin/getdesc?name=ig). Transmembrane protein-related sequences can also possess or interact with ion_trans domains, which are polypeptides characterized by six transmembrane helices, and which transport ions across membranes (http://pfam.wustl.edu/cgi-bin/getdesc?name=ion_trans). Proteins in this family can demonstrate specificity for particular ions, e.g., sodium, potassium, and calcium. Transmembrane protein-related sequences can also possess or interact with integrase core domains, which mediate the integration of a DNA copy of a viral genome into a host chromosome; e.g., HIV integrase catalyses the incorporation of virally derived DNA into the human genome, presenting a target for the development of new therapeutics for the treatment of AIDS (http://pfam.wustl.edu/cgi-bin/getdesc?name=rve). Transmembrane protein-related sequences can also possess or interact with domains designated as differentially expressed in neoplastic vs. normal cells "DENN" domains, which are involved in signal transduction. Characteristically, these domains are found in protein components of signaling pathways that utilize rab proteins or mitogen-activated protein (MAP) kinases (http://pfam.wustl.edu/cgi-bin/getdesc?name=DENN).

[033] Transmembrane protein-related sequences can also possess or interact with acyl coA binding protein (ACBP) domains, which are protein domains that bind medium- and long-chain acyl-CoA esters with high affinity (http://pfam.wustl.edu/cgi-bin/getdesc?name=ACBP). Membrane-related sequences also possess or interact with SPFH domain/band 7 family (Band_7) domain, which are protein domains that include a transmembrane segment, and regulate cation conductivity (http://pfam.wustl.edu/cgi-bin/getdesc?name=Band_7).

[034] Transmembrane proteins that are differentially expressed on the surface of cancer cells, particularly those that are differentially expressed on the surface of cancer cells but not on the surface of normal tissues, such as heart and lung, are desirable targets for production of antibodies, e.g., diagnostic antibodies or therapeutic antibodies, such as antibodies that mediate ADCC or CDC to effect tumor cell killing.

[035] Transmembrane proteins with extracellular fragments that can be cleaved can be useful as secreted proteins to effect ligand/receptor binding so as to mediate intracellular responses, such as signal transduction. Transmembrane proteins that act as receptors, and possess a ligand binding extracellular portion exposed on a cell surface and an intracellular portion that interacts with other cellular components upon activation can be also be useful as transmembrane proteins to mediate intracellular responses, such as signal transduction.

Kinase-Related Sequences

[036] A kinase is an enzyme that catalyzes the transfer of phosphate groups from phosphate donors to acceptor substrates. Kinase substrates include, but are not limited to, proteins and lipids. Sequences of the invention that phosphorylate protein substrates are designated "Pkinases." Examples of kinase-related sequences include calcium, calmodulin-dependent protein kinase II, myosin light chain kinase, and phosphatidlyinositol kinase.

[037] Kinases and phosphatases are counteracting: kinases add phosphate groups and phosphatases liberate phosphate groups. The counteracting activities of kinases and phosphatases provide cells with a "switch" that can turn on or turn off the function of various proteins. The activity of any protein regulated by phosphorylation depends on the balance, at any given time, between the activities of the kinase(s) that phosphorylate it, and the phosphatase(s) that dephosphorylate it. Phosphorylation plays a important role in intercellular communication during development, homeostasis, and the function of major bodily systems, including the immune system.

[038] In conjunction with phosphatases, kinases control such diverse and essential cellular processes as transcription, cell division, cell cycle progression, differentiation, cytoskeletal function, apoptosis, receptor function, learning and memory, hematopoeisis, fertilization, neural transmission, muscle contraction, non-muscle motor function, glycogen metabolism, and hormone secretion.

[039] Most kinases act within a network of kinases and other signaling effectors, and are modulated by autophosphorylation and phosphorylation by other kinases (Manning et al., 2002). Intracellular signaling involves a multitude of diverse mechanisms that combine to modulate the activity of individual proteins in response to different biological inputs.

[040] Defects in cell signal transduction pathways are responsible for a number of disorders, including the majority of cancers, immune disorders, and many inflammatory conditions, including, but not limited to, Crohn's disease (Geffen and Man, 2002; Van Den Blink et al., 2002; Lodish 1999). Over-expression and/or structural alteration of kinases, for example, receptor tyrosine kinase family members, is often associated with human cancers. For example, tumor cells are known to use receptor tyrosine kinases in transduction pathways to achieve tumor growth, angiogenesis and metastasis. Therefore, receptor tyrosine kinases represent pivotal targets in cancer therapy. A number of small molecule receptor tyrosine kinase inhibitors have been synthesized, are in clinical trials, are being analyzed in animal models, or have been marketed. Inhibitory mechanisms include ligand-dependent down regulation, e.g., by the adaptor Cbl (Brunelleschi et al., 2002).

[041] Kinase-related sequences can possess or interact with protein kinase (pkinase) domains, which share a conserved catalytic core common in serine/threonine and tyrosine protein kinases (http://pfam.wustl.edu/cgi-bin/getdesc?name=pkinase). Kinase-related sequences can also possess or interact with A-kinase anchoring protein 95 (AKAP95) domains, which comprise two zinc fingers, and have been implicated in chromosome condensation (http://pfam.wustl.edu/cgi-bin/getdesc?name=AKAP95). Kinase-related sequences can also possess or interact with inositol 1,3,4,-trisphosphate 5/6 kinase (Ins134_P3_kin) domains, which mediate the function of inositol 1.3.4-trisphosphate, a branch point in inositol phosphate metabolism (http://pfam.wustl.edu/cgi-bin/getdesc?name= Ins134_P3_kin).

[042] Kinases, by virtue of their participation in many and varied intracellular activities, are useful as targets of therapeutic intervention such as, for example, in cancer and inflammation. Cells transfected with cDNA encoding a kinase can be used in screening for small molecule agonists or antagonists, for example.

Ligase-Related Sequences

[043] Ligases are enzymes that join together, or ligate, two molecules. Ligase substrates include nucleic acids and proteins. For example, DNA ligases link

two DNA molecules together; they play a role in DNA repair and replication. DNA ligases also are involved in the rearrangement of immunoglobulin gene segments, such as those responsible for the generation of antibody diversity. Examples of protein ligases include ubiquitin protein ligases, which add an ubiquitin molecule to an amino acid residue, typically as part of a peptide or polypeptide. Examples of nucleic acid ligases include DNA ligase I, DNA ligase III alpha, and T4 RNA ligase 2.

[044] Ligases are also involved in cellular regulatory processes. For example, glutamate-cysteine ligase (GCL) is the first and rate-limiting enzyme involved in the biosynthesis of glutathione. Polymorphisms of human GCL account for differences in sensitivity to environmental toxicants and chemotherapeutic agents in human cancer cell lines (Walsh et al., 2001). Also by way of example, glutamate-ammonia ligase, or glutamine synthetase (GS), is expressed at a higher than normal level in human primary liver cancer, and may be involved in hepatocyte transformation (Christa et al., 1994).

[045] Ligase-related sequences can possess or interact with ATP dependent DNA ligase (DNA_ligase) domains, which can join two DNA fragments by catalyzing the formation of an internucleotide ester bond between a phosphate and a deoxyribose (http://pfam.wustl. edu/cgi-bin/getdesc?name= DNA_ligase). Ligase-related sequences can also possess or interact with glutamate-cysteine ligase (GCS) domains, which catalyze the rate-limiting step in the biosynthesis of glutathione. (http://pfam.wustl.edu/cgi-bin/getdesc?name=GCS). Ligase-related sequences can also possess or interact with 2',5' RNA ligase (2_5_ligase) domains, which ligate tRNA half molecules containing 2',3'-cyclic phosphate and 5' hydroxyl terminal to products containing a 2'5' phosphodiester linkage (http://pfam.wustl.edu/cgi-bin/getdesc?name=2_5_ligase).

[046] Like kinases, ligases are also useful as targets for identification of agonists and antagonists, such as small molecule drugs.

Receptor-Related Sequences (Including Nuclear Hormone and T-Cell Receptors)

[047] A receptor is a polypeptide that binds to a specific signaling molecule and initiates a cellular response. Receptors can be present on the cell surface or inside the cell. Example of receptor types include G-protein-linked receptors, ion channel-linked receptors, enzyme-linked receptors, T-cell receptors, thyroid hormone receptors, retinoid receptors, nuclear hormone receptors, and the

related category of steroid hormone receptors, e.g., cortisol receptors (Alberts et al., 1994).

[048] G-protein-linked receptors transduce extracellular signals into intracellular responses by interacting with guanine nucleotide binding proteins. The same ligand can activate many different G-protein-linked receptors. G-protein-linked receptors mediate cellular responses to a diverse range of signaling molecules, including hormones, neurotransmitters, and local mediators, which are varied in structure and function, and encompass proteins and small peptides, as well as amino acids and their derivatives, and fatty acids and their derivatives. Many signaling molecules are active at low concentrations, and their receptors often bind with high affinity. Examples of G-protein-linked receptors include, but are not limited to, rhodopsins, olfactory receptors, and β-adrenergic receptors.

[049] Ion channel-linked receptors are involved in synaptic signaling. These receptors regulate ion channels, to which they are linked. Some respond to signals from neurotransmitters, e.g., acetylcholine, serotonin, GABA, and glycine. A common mechanism of action for ion channel-linked receptors is to transiently open or close their respective ion channel, transiently changing the permeability of the membrane in which they reside to a specific ion or ions.

[050] Enzyme-linked receptors can be linked to enzymes or can function as enzymes. Their ligand binding site is commonly on one side of the membrane, e.g., an extracellular domain, and the catalytic site is on the other, e.g., a cytoplasmic domain. Transmembrane tyrosine-specific protein kinase receptors for growth and differentiation factors are enzyme-linked receptors; examples include receptors for epidermal growth factor (EGF), platelet-derived growth factor (PDGF), fibroblast growth factors (FGFs), hepatocyte growth factors (HGF), insulin, insulin like growth factor-1 (IGF-1), nerve growth factor (NGF), vascular endothelial growth factor (VEGF), and macrophage colony stimulating factor (M-CSF).

[051] Nuclear hormone receptors generally function by crossing the plasma membrane of target cells and binding to intracellular protein ligands. Ligand binding activates these receptors in some instances, exposing a DNA binding domain which regulates the transcription of specific genes. Generally, nuclear hormone receptors bind to specific DNA sequences adjacent to or in the vicinity of the genes regulated by their ligand. A host of cell type-specific regulatory proteins can

collaborate with the nuclear hormone receptor to influence the transcription of specific genes or sets of genes (Alberts et al., 1994). Examples of nuclear hormone receptors include estrogen-related receptors, such as hERR1, which modulates the estrogen receptor-mediated response of the lactoferrin gene promoter (Yang et al., 1996), and is a transcriptional regulator of the human medium chain acyl coenzyme A dehydrogenase gene (Sladek et al., 1997). Examples of nuclear hormone receptors also include photoreceptor-specific nuclear receptors, such as NR2E3, which are part of a large family of nuclear receptor transcription factors involved in signaling pathways. NR2E3 plays a role in cone function and human retinal photoreceptor differentiation and degeneration (Milam et al., 2002; Kobayashi et al., 1999).

[052] T-cell receptors are membrane proteins comprised of two disulfide-linked polypeptide chains, each with two immunoglobulin-like domains. They display a similarity to antibodies in that they have a variable amino-terminal region and a constant carboxyl-terminal region which is coded for by variable, joining, and constant region genes (Wei et al., 1997; Alberts et al., 1994). Rearrangement of T-cell receptor genes have been associated with human T-cell leukemias (Fisch et al., 1993).

[053] Receptors are involved in cellular processes that regulate growth and differentiation. Their dysregulation can lead to hyperproliferative conditions, and they are common therapeutic targets. For example, the EGF receptor is aberrantly activated in neoplasia, especially in tumors of epithelial origin. EGF receptor antagonists can successfully treat some of these tumors, either alone or in combination with chemotherapy or ionizing radiation (Kari et al., 2003). The progesterone receptor, an intracellular steroid hormone receptor, plays a role in the development and function of the mammary gland, the uterus, and the ovary. Mutation or aberrant expression of the progesterone receptor, or its regulatory molecules, can affect its normal function and lead to cancer (Gao and Nawaz, 2002).

[054] Receptors are also involved in cellular processes that regulate inflammation and immunity. For example, members of the type 1 interleukin-1 receptor family mediate immune and inflammatory responses, and function in host defense. (O'Neill, 2002). Their activation can lead to the activation of signaling cascades, e.g., pathways involving transcription factors and protein kinases, resulting in an inflammatory response (O'Neill, 2002). Another mechanism by which receptors regulate inflammation and immunity is by their selective expression, at discrete stages

of differentiation, by cells involved in the inflammatory response. For example, expression of the triggering receptor expressed on myeloid cells (TREM-1) and the myeloid DAP12-associating lectin (MDL-1) are correlated with myelomonocytic differentiation. These receptors are more highly expressed in differentiated cells, are involved in monocyte activation and the inflammatory response, and are expressed at a lower level in malignant compared to normal cells (Gingras et al., 2002).

[055] Receptor-related sequences can possess or interact with seven transmembrane receptor (7tm_1) domains, which are protein domains with a structural framework comprising seven transmembrane helices found in receptors, e.g., receptors in the rhodopsin family with a wide range of functions, activated by ligands that vary widely in structure and character (http://pfam.wustl.edu/cgi-bin/getdesc?name=7tm_1). Receptor-related sequences can also possess or interact with L1 transposable element (transposase_22) domains, some of which have been characterized to exhibit reverse transcriptase activity, and some of which are capable of retrotransposition. Receptor-related sequences can also possess or interact with a SH2 domain, which is a protein domain of about 100 amino acid residues found in many intracellular signal-transducing proteins, that can regulate intracellular signaling cascades by interacting with phosphotyrosine-containing target peptides in a sequence-specific and phosphorylation-dependent manner (http://pfam.wustl.edu/cgi-bin/getdesc?name=SH2). Receptor-related sequences can also possess or interact with LDL receptor domains, e.g., the low-density lipoprotein receptor repeat class B (Ldl_recept_b) domain, which comprises a conserved YWTD motif in multiple tandem repeats (http://pfam.wustl.edu/ cgi-bin/getdesc?name=ldl_recept_b). Receptor-related sequences can also possess or interact with ribosomal L10 (Ribosomal_L10e) domains, which are protein domains commonly found in the large ribosomal subunit (http://pfam.wustl.edu/cgi-bin/getdesc?name=Ribosomal_L10e).

[056] Receptor-related sequences can possess or interact with zinc finger C4 type domains, which are DNA binding domains of nuclear hormone receptors that share a conserved cysteine-rich region of approximately 65 amino acids and regulate such diverse biological processes as pattern formation, cellular differentiation, and homeostasis (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00105). Receptor-related sequences can also possess or interact with a ligand binding domain of nuclear hormone receptors (hormone_rec), which are helical domains involved in the regulation of eukaryotic gene expression, cellular

proliferation, and differentiation in target tissues (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00104). Receptor-related sequences can also possess or interact with Mov34 domains, which are regulatory subunits of the proteasome found in some regulators of transcription factors (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF01398). Receptor-related sequences can also possess or interact with immunoglobulin domains, which are described above.

[057] Receptors, and fragments of receptors can be used as therapeutics. For example, a ligand-binding portion, an effector-binding portion, and a kinase or phosphatase domain or consensus sequence can comprise fragments that can function as agonists or antagonists enhance or reduce, e.g., ligand binding to the natural receptors, or effector function by the natural receptors.

Phosphatase-Related Sequences

[058] A phosphatase, as indicated above, is an enzyme that catalyses the hydrolysis of esters of phosphoric acid. Its substrates include, but are not limited to, nucleic acids, proteins, and lipids. Together with kinases, phosphatases are active in a broad range of cellular functions, including transcription, cell division, cell-cycle progression, intermediate cellular metabolism, glycogen metabolism, lipogenesis and lipolysis, maintenance of electrochemical gradients, neuronal function, immune responses, intracellular vesicular transport, cytoskeletal function, sperm motility, and skeletal, cardiac, and smooth muscle function (Oliver and Shenolikar, 1998).

[059] Disruption in these functions may lead to disorders. For example, as noted above, phosphatases regulate pathways of cell growth and programmed cell death; disruptions in these pathways can lead to abnormal cell growth, such as that which occurs in cancer. Mutations in serine/threonine protein phosphatase 2A (PP2A), a multifunctional regulator of cell growth and function, are associated with the increased growth of tumor cells (Schonthal, 2001). The tumor suppressor "phosphatase and tensin-homology deleted on chromosome 10" (PTEN) gene encodes PIP_3, a lipid phosphatase that dephosphorylates phosphatidlyinositol, thus countering the action of the oncogenes PI_3-kinase and Akt, which promote cell survival. PTEN has been identified as a tumor suppressor; it is deleted in multiple types of advanced human cancers.

[060] Also as noted above, phosphatases regulate pathways that control immune function. For example, the CD45 phosphotyrosine phosphatase is one of the most abundant glycoproteins expressed on immune cells, and regulates T-cell

signaling and development (Alexander, 2000). In addition, the serine/threonine phosphatase calcineurin plays a central role in lymphocyte activation, among other important and wide-ranging cellular functions (Baksh and Burakoff, 2000). Certain compounds, specifically, cyclosporine and FK-506 (Tacrolimus), have been found to inhibit the phosphatase activity of calcineurin, thereby suppressing the production of IL-2 and other cytokines. In addition, these compounds have recently been found to block the JNK and p38 signaling pathways triggered by antigen recognition in T-cells. Finally, phosphatase inhibitors have proven to be valuable as immune suppressant drugs, and those in the field believe that modulators of phosphatase activity promise to be important immunoregulatory compounds (Allison, 2000).

[061] Phosphatase-related sequences can possess or interact with protein phosphatase 2C (PP2C) domains, which display Mn^{++} or Mg^{++} dependent protein serine/threonine phosphatase activity (http://pfam.wustl.edu/cgi-bin/getdesc?name=PP2C). Phosphatase-related sequences can also possess or interact with protein-tyrosine phosphatase (Y_phosphatase) domains, which catalyze the removal of a phosphate group attached to a tyrosine residue (http://pfam.wustl.edu/cgi-bin/getdesc?name=Y_phosphatase). Phosphatase-related sequences can also possess or interact with protein phosphatase inhibitor 1/DARPP-32 (DARPP-32) domains, which inhibit protein phosphatases, and play a role in regulating neurotransmitter pathways, receptors, and ion channels (http://pfam.wustl.edu/cgi-bin/getdesc?name=DARPP-32).

[062] Like kinases, phosphatases can be used as targets for therapeutic intervention, in cell-free or cell-based assays, for example, in screening for drugs, including small molecule drugs.

Protease-Related Sequences

[063] Proteases, also known as endopeptidases, are enzymes that cleave polypeptide chains by hydrolyzing peptide bonds at positions within the amino acid chain. Different proteases recognize different polypeptide sequences. Endopeptidase substrate specificities vary from broad to narrow; for example, subtilisins are relatively non-specific, and can cleave polypeptide chains with a wide variety of amino acid sequences, whereas thrombin is more specific and can only cleave polypeptide chains with an arginine residue on the carboxyl side of the susceptible peptide bond and glycine on the amino side. Additional examples of protease-related

sequences include collagenases, trypsin, and damage-induced neuronal endopeptidase (Kiryu-Seo et al., 2000).

[064] Proteases mediate the continuous remodeling of living tissues. For example, the extracellular matrix, a tissue skeleton that mediates communication among cells, and influences the structure and function of associated tissues and organs, is continuously remodeled. A strictly controlled balance is maintained between breakdown of the extracellular matrix by proteases and reconstruction of the extracellular matrix. This continued matrix remodeling is a dynamic process that shapes the structure and function of tissues and organs (Wojtowicz-Praga, 1999).

[065] Defects in protease function are responsible for a number of disorders, including cancer and other hyperproliferative disorders. Proteases are involved in the pathogenesis of such disorders both by virtue of their involvement in programmed cell death and tumor invasion and metastasis (Los et al., 2003; Stetler-Stevenson et al., 1993). Detection of the presence or characteristics of proteases can be used to screen for and diagnose prostate cancer (Karanazanashvili and Abrahamsson, 2003). Proteases are also involved in the pathogenesis of inflammatory and arthritic diseases, such as pancreatitis, osteoarthritis, and rheumatoid arthritis (Pfutzer and Whitcomb, 2001; Martel-Pelleteir et al., 2001; Lerch and Gorelick, 2000).

[066] Protease-related sequences possess or interact with a variety of different protease domains, including domains belonging to the cysteine protease family, the serine protease family, and the metalloproteinase family (http://pfam.wustl.edu/cgi-bin/text search?terms=endopeptidase&search_what= all&sections =DE&sections=CC&size=10).

Phosphodiesterase-Related Sequences

[067] Phosphodiesterases are enzymes that cleave phosphodiester bonds, i.e., bonds formed by two hydroxyl groups in an ester linkage to the same phosphate group, such as those between adjacent RNA or DNA nucleotides. Phosphodiesterases are found in both soluble and membrane-associated forms. Most phosphodiesterases act within a network of signal transduction molecules and other signaling effectors, and are modulated by components of these pathways. Phosphodiesterases regulate the metabolism and synthesis of cyclic nucleotides in signal-transduction pathways. They hydrolyze cAMP and cGMP, molecules that play an important and widespread role in signal transduction. Phosphodiesterases also

repair damage to nucleic acids. Some phosphodiesterases are regulated primarily by calcium and calmodulin, others are regulated primarily by cGMP. They differ in their sensitivity to individual inhibitors, but all share a homologous catalytic region (Siegel, et al., 1999).

[068] Examples of phosphodiesterases include nucleotide pyrophosphatases (NPP) and plasma membrane glycoprotein PC-1, which are present in elevated levels in the fibroblasts of patients with Lowe's syndrome (Funakoshi et al., 1992). Another example of a phosphodiesterase is myomegalin-like protein, which is expressed at high levels in the nucleus and cytoplasm of heart and skeletal muscle (Soejima et al., 2001). Phosphodiesterases have demonstrated promise in cancer chemotherapy, analgesia, the treatment of Parkinson's disease, and the treatment of learning and memory disorders (Weishaar, et al., 1985).

[069] Phosphodiesterase-related sequences can possess or interact with type I phosphodiesterase/nucleotide pyrophosphatase (phosphodiest) domains, which catalyze the cleavage of phosphodiester and phosphosulfate bonds (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF01663). Phosphodiesterase-related sequences can also possess or interact with 3'5'-cyclic nucleotide phosphodiesterase (PDEase) domains, which are involved in signal transduction (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00233).

[070] Phosphodiesterases (PDEs) are also useful as targets for therapeutic intervention, for example, for identification of agonists or antagonists, such as in the screening of small molecule inhibitors. A well known PDE-5 inhibitor, sildenafil citrate (Viagra®) is used for treatment of erectile dysfunction (Brock, 2000). The mechanism of action involves inhibition of PDE-5 enzyme and resulting increase in cyclic guanosine monophosphate (cGMP) and smooth muscle relaxation in the penis (Rosen and McKenna, 2002). Such inhibitors may also find use for treatment of severe pulmonary arterial hypertension. (Ghofrani et al., 2003).

Kinesin-Related Sequences

[071] Cells transport proteins and organelles in an orderly and regulated manner along cytoskeletal filaments. Molecular motor proteins, such as kinesins, can carry such cargo along the cytoskeletal filaments to specific destinations, in a highly regulated manner. Exemplary membrane-bound cargoes include mitochondria, lysosomes, endoplasmic reticulum, and axonal vesicles (Vale,

2003). Kinesins also transport nonmembranous cargo, such as mRNAs, tubulin monomers, and intermediate filaments (Vale, 2003).

[072] Kinesins, e.g., KIF11, function in the cell division process (Miki et al., 2001). In the nucleus, kinesins are necessary to establish spindle bipolarity, position chromosomes on metaphase plates, and maintain forces in the spindle. Several members of the kinesin family are associated with the chromosomes, and are likely to perform a role in mitotic chromosome movement (Miki et al., 2001). For example, the C-terminal kinesin KIFC1 is involved in the processes of meiosis, mitosis, and karyogamy (Miki et al., 2001). The kinesin GAKIN binds to the human analog of the Drosophila Discs Large tumor suppressor protein (hDlg), a membrane associated guanylate kinase (Hanada, 2000). GAKIN undergoes translocation in T-lymphocytes upon their cellular activation (Hanada, 2000). The GAKIN/hDlg complex is also hypothesized to play a role in cell division (Hanada, 2000). Thus, the kinesin GAKIN plays a role in cell proliferation and T-cell mediated immune function.

[073] Kinesin-mediated intracellular transport is also implicated in as a mechanism of tumorigenesis. For example, kinesin transports the tumor suppressor adenomatous polyposis colon protein (APC) (Jimbo et al., 2002). The APC gene is mutated in both sporadic and familial colorectal tumors. The APC protein interacts with the microtubule plus-end-directed kinesin proteins KIF3A and KIF3B through an association with the kinesin superfamily-associated protein 3 (KAP3). Normally, the APC tumor suppressor is transported to its correct intracellular location at the tips of membrane protrusions. Mutant APCs derived from cancer cells, however, are unable to undergo kinesin-mediated transport, and do not accumulate with normal efficiency in clusters in the membrane protrusions, and thereby can not function efficiently as tumor suppressors.

[074] In view of the connection to cancer, investigators have sought small molecules to inhibit specific molecular motors in cells, such as the mitotic kinesin Eg5/Ksp (Mayer, 1999). In addition, others have found small molecule inhibitors of Eg5/Kap with low nanomolar affinity have anti-tumor activity, and one such agent has entered clinical phase I trials (Vale, 2003).

[075] In another arena, it has been proposed that impairing motor-driven delivery of MHC peptide complexes to the surface of dendritic cells could provide immunomodulation. Additionally, inhibiting the cell surface delivery of

cytotoxic granules in T cells could help provide immunosuppressive therapy (Vale, 2003).

[076] Kinesin-related sequences can possess or interact with kinesin motor (kinesin) domains, which hydrolyze ATP and bind to microtubules to produce a motor-active force that transports intracellular vesicles and organelles (http://pfam.wustl.edu/cgi-bin/getdesc?name=kinesin). Kinesin-related sequences can also possess or interact with kinesin-associated protein (KAP) domains, which are non-motive domains that form a complex with kinesin (http://pfam.wustl.edu/cgi-bin/getdesc?name=KAP). Kinesin-related sequences can also possess or interact with MyTH4 domains, which are present in the tail of the motor ATPase proteins kinesin and myosin (http://pfam.wustl.edu/cgi-bin/getdesc?name=MyTH4).

[077] Kinesins, like kinases, are useful as targets for therapeutic intervention, for example, in screening for small molecule inhibitors for the treatment of cancer.

Immunoglobulin-Related Sequences

[078] An immunoglobulin is an antibody molecule, and is typically composed of heavy and light chains, each of which have constant regions that display similarity with other immunoglobulin molecules and variable regions that convey specificity to particular antigens. Most immunoglobulins can be assigned to classes, e.g., IgG, IgM, IgA, IgE, and IgD, based on antigenic determinants in the heavy chain constant region; each class plays a different role in the immune response.

[079] Immunoglobulins are characterized by a structural motif, the immunoglobulin (ig) domain, which is approximately one hundred amino acids long, is involved in protein-protein and protein-ligand interactions, and includes a conserved intradomain disulfide bond (http://pfam.wustl.edu/cgi-bin/getdesc?name=ig). It is one of the most common domains found among all known proteins, and is present in hundreds of proteins with diverse functions. Proteins with the ig domain comprise the immunoglobulin superfamily; members include antibodies, T-cell receptors, major histocomptability proteins, the CD4, CD8, and CD28 co-receptors, most of the invariant polypeptide chains associated with B and T cell receptors, leukocyte F_c receptors, the giant muscle kinase titin, and receptor tyrosine kinases (Janeway et al., 2001; Alberts, et al., 1994).

[080] Polypeptides with immunoglobulin-like domains can be markers for specific types of tissues and tumors. For example, a 43-kDa protein membrane

antigen with two immunoglobulin-like domains in its extracellular region is expressed in normal human colonic and small bowel epithelium and > 95% of human colon cancers, but absent from most other human tissues and tumor types (Heath et al., 1997).

[081] Polypeptides with immunoglobulin-like domains are also involved in inflammation. For example, myelin oligodendrocyte glycoprotein, a myelin-specific protein found in the central nervous system, specifically binds to and activates complement, an effector of the immune system, via its extracellular immunoglobulin-like domain. By virtue of providing the means for an interaction between myelin and the complement component of the immune response, myelin oligodendrocyte glycoprotein is a modulator of central nervous system inflammation and has been predicted by those in the field to be relevant to the pathogenesis of demyelinating diseases such as multiple sclerosis (Johns and Barnard, 1997).

[082] Immunoglobulin-related sequences can also possess or interact with leucine-rich repeat domains, which are involved in protein-protein interactions, and are used in molecular recognition processes as diverse as signal transduction, cell adhesion, cell development, DNA repair and RNA processing (http://pfam.wustl.edu/cgi-bin/getdesc?name =LRRNT). Immunoglobulin-related sequences can also possess or interact with fibronectin type III repeat (fn3) domains (http://pfam.wustl.edu/cgi-bin/getdesc?name=fn3), which contain binding sites for DNA and heparin. Immunoglobulin-related sequences can also possess or interact with WASp Homology domain 1 (WH1), which can bind the metabotropic glutamate receptors mGluR1alpha and mGluR5 (http://pfam.wustl.edu/cgi-bin/getdesc?name=WH1).

Glycosylphosphatidylinositol Anchor-Related Sequences

[083] Glycosylphosphatidylinositol (GPI) anchor proteins are synthesized as single membrane proteins; the transmembrane segment is cleaved away in the endoplasmic reticulum, where a GPI membrane anchor is added. The resulting protein is bound to the non-cytoplasmic, i.e., either extracellular or luminal, side of the membrane by the GPI anchor. GPI anchor proteins can be dissociated from the membrane by phosphatidylinositol-inositol-specific phospholipase C (Alberts et al., 1994). Examples of GPI-anchor proteins include prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin (Vainberg et al.,

1998), and carboxypeptidase M, which is associated with the differentiation of monocytes to macrophages (Rehli et al., 1995).

[084] GPI anchor protein-related sequences can possess or interact with KE2 domains, which may contain a DNA binding leucine zipper motif(http://www.sanger.ac.uk /cgi-bin/Pfam/getacc?PF01920). GPI anchor protein-related sequences can also possess or interact with zinc carboxypeptidase (Zn_carbOpept) domains, which include carboxypeptidase H regulatory domains and carboxypeptidase A digestive domains (http://www.sanger.ac.uk/cgi-bin/Pfam/getacc?PF00246).

Other Polypeptide-Related Sequences

Activator-Related Sequences

[085] An activator is a molecule or collection of molecules that positively modulates the activity of a regulatory protein, or that binds to DNA and regulates one or more genes by increasing the rate of transcription. Regulatory protein activators contribute to an increase in protein activity. Transcriptional activators provide a positive control over gene transcription; for example, they can sense the internal condition of the cell and bind to a sequence of DNA near a target promoter, resulting in the transcription of an appropriate gene. Examples of activator-related sequences include template-activating factors, bacterial catabolite activators, and the coenzyme thiamine pyrophosphatase. Activator-related sequences, e.g., factors that influence viral replication and transcription, can be encoded by oncogenes (Nagata et al., 1995).

[086] Activator-related sequences can possess or interact with SH2 domains, which are protein domains of about 100 amino acid residues found in many signal-transducing proteins. SH2 domains can regulate signaling cascades, e.g., by interacting with phosphotyrosine-containing target peptides in a sequence-specific and phosphorylation-dependent manner (http://pfam.wustl.edu/cgi-bin/getdesc?name=SH2). Activator-related sequences also possess or interact with nucleosome assembly protein (NAP) domains, which regulate gene expression, and are accessible to histones (http://pfam.wustl.edu/cgi- bin/getdesc?name=NAP).

Adaptor-Related Sequences

[087] Adaptors are proteins involved in the process of capturing specific cargo molecules into membrane-bound vesicles for transport through the cell. Different adaptors recognize different receptors for cargo molecules, and also recognize different vesicle coat proteins, accounting, in part, for the specificity of the

content of intracellular vesicles bound to specific destinations within the cell (Kirsch et al., 1999). Examples of adaptor-related sequences include adaptins, clathrins, adaptor-related protein complex subunits, and Cas ligand with multiple Src homology 3 domains (CMS) adaptors.

[088] Adaptor-related sequences can possess or interact with src homology 3 (SH3) domains, which are small protein modules of approximately 50 amino acid residues found in a variety of intracellular or membrane-associated proteins. SH3 domains are often indicative of a protein involved in signal transduction events related to cytoskeletal organization. (http://pfam.wustl.edu/cgi-bin/getdesc?name=SH3). Adaptor-related sequences also possess or interact with the adaptin N-terminal (Adaptin_N) protein domain, which is found in the N terminal region of various adaptor protein complexes. The N-terminal region of adaptor proteins is relatively constant in comparison to the C-terminal (http://pfam.wustl.edu/cgi-bin/getdesc?name=Adaptin_N).

Adhesion Molecule-Related Sequences

[089] Adhesion molecules are molecules that mediate the adhesion of cells with other cells, and with the extracellular matrix. Examples of adhesion molecules include members of the immunoglobulin superfamily, integrins, cadherins, selectins, and transmembrane proteoglycans. The adhesion molecule carcinoembryonic antigen (CEA) is present nearly exclusively on cancer cells, and is expressed on the cell surface of approximately 80% of all solid cancerous tumors (Berinstein et al., 2002).

[090] Adhesion molecule-related sequences can possess or interact with the immunoglobulin (ig) domain, which are described above. Adhesion molecule-related sequences can also possess or interact with integrin alpha cytoplasmic region (integrin_A) domains, which comprise the short, intracellular region of the integrin alpha chain http://pfam.wustl.edu/cgi-bin/getdesc?name=integrin_A).

Antigen-Related Sequences

[091] An antigen is a molecule that provokes an immune response; they include both foreign antigens and autoantigens. Antigens can be expressed in a tissue-specific manner and their expression can be developmentally regulated. For example, the heat stable antigen HSA is expressed in both a tissue-specific manner, i.e., it is restricted to hematopoeitic cells, and a developmentally-regulated manner, i.e., it is more highly expressed in immature precursor cells than in terminally

differentiated cells (Wenger et al., 1993). Antigens can be expressed on the cell surface or inside the cell, e.g., in the nucleus or on intermediate filaments. Antigen-related sequences include sequences related to tumor antigens, which are expressed exclusively in tumor cells, or in greater amounts in tumor cells than in normal cells. Tumor antigens can be transmembrane proteins, with one or more transmembrane domains (Li et al., 1996; Linnenbach, et al., 1993).

[092] Autoantigens, which are components of the body that provoke an immune response, are involved in the pathogenesis of autoimmune disease. Autoantigens can be either selectively or ubiquitously expressed among cell and tissue types. They can be localized to any region of the cell, including the nucleus, nucleolus, nuclear envelope, and intermediate filaments (Racevskis et al., 1996). For example, pancreatic islet cell antigens are involved in the autoimmune pathogenesis of diabetes, and thyroid antigens are involved in autoimmune thyroid disease.

[093] Antigen-related sequences can possess or interact with the ICAp69 domain, which is characterized by a 69 kDa pancreatic islet cell autoantigen present in autoimmune (insulin-dependent) diabetes mellitus (http://pfam.wustl.edu/cgi-bin/getdesc?name=ICA69). Antigen-related sequences can also possess or interact with the Ku70/Ku80 C-terminal arm (Ku_C) or Ku70/Ku80 N-terminal alpha/beta (Ku_N) domains, which belong to the Ku family of peptides (http://pfam.wustl. edu/cgi-bin/getdesc?name=Ku_C; http://pfam.wustl. edu/cgi-bin/getdesc? name=Ku_N). Ku, an antigen associated with autoimmune disease, normally functions to bind DNA double-strand breaks and facilitate DNA repair, but induces autoimmunity under pathological conditions. Antigen-related sequences can also possess or interact with the bZIP transcription factor (bZIP) domain, which comprises a basic region and a leucine zipper region (http://pfam.wustl.edu/cgi-bin/getdesc? name=bZIP). Antigen-related sequences can possess or interact with YT521-B-like (YTH) domains, which comprise YT521-B, a tyrosine-phosphorylated nuclear protein domain that modulates alternative RNA splice site selection, and interacts with other nuclear proteins, e.g., scaffold attachment factor B, and Sam68, a 68-kDa substrate associated with Src during mitosis (http://pfam.wustl.edu/ cgi-bin/getdesc?name= YTH).

ATPase-Related Sequences

[094] ATPases are enzymes that use the energy of ATP hydrolysis to move ions or small molecules across a membrane against a chemical concentration

gradient or electrical potential. For example, ATPases can maintain low intracellular calcium and sodium ion concentrations, and generate a low pH inside lysosomes, plant-cell vacuoles, and the lumen of the stomach. Vacuolar ATPases are ATP-dependent proton pumps that create pH gradients by transporting protons across membranes, while coupling the energy produced in the conversion of ATP to ADP with proton transport (Forgac, 1999). They can acidify or alkalinize cells, organelles, and extracellular compartments, and create voltage gradients that drive the secretion or absorption of ions and fluids (Wieczorek et al. 1999). Examples of ATPase-related sequences include proton transporters, glucose transporters, multidrug resistance factors, calcium ATPases, and porins.

[095] ATPase-related sequences can possess or interact with ATP synthase F/14-kDa subunit (ATP-synt-F) domains, which correspond to a 14-kDa subunit in the peripheral catalytic part of vacuolar ATPases (http://pfam.wustl.edu/cgi-bin/getdesc?name=ATP-synt_F). ATPase-related sequences can also possess or interact with vacuolar (H^+)-ATPase C, D, G, and H subunit (V-ATPase) domains, which are membrane-attached sequences that generate an acidic environment (http://pfam.wustl.edu/cgi-bin/getdesc?name=V-ATPase_G).

ATP-Related Sequences

[096] Adenosine trisphosphate (ATP) is a nucleotide comprising an adenine, a ribose, and a trisphosphate unit. The trisphosphate unit contains two phosphoanhydride bonds that confer an energy-rich property to ATP. The free energy liberated in the hydrolysis of one or both of these bonds can drive reactions that require an input of free energy. A wide range of physiological and pathological processes are driven by the energy of ATP, including cellular movement, the synthesis of biomolecules from precursors, muscle contraction, ciliary and flagellar function, intermediary metabolism, glycolysis, fatty acid oxidation, oxidative phosphorylation, and membrane transport (Ku et al., 1990). Examples of ATP-related sequences include ATPases, ATP synthases, ATP carrier proteins, and myosin.

[097] ATP-related sequences can possess or interact with ATP-synthase subunit C protein domains (ATP-synt_C), which are protein domains that consist of two long terminal hydrophobic regions, and are implicated in the proton-conducting activity of ATPases (http://pfam.wustl.edu/cgi-bin/getdesc?name=ATP-synt_C). ATP-related sequences can also possess or interact with mitochondrial carrier protein (mito_carr) domains, which are involved in energy transfer across the

inner mitochondrial membrane (http://pfam.wustl.edu/cgi-bin/getdesc? name= mito_carr).

Binding Protein-Related Sequences

[098] A binding protein is a protein that binds to another molecule with specificity. Binding proteins can be involved in building macromolecular structures, e.g., in cytoskeletal assembly or scaffolding (Machesky et al., 1997). Proteins often exist in the cell in complexes with other proteins, nucleic acids, lipids, and/or small molecules. For example, steroid receptors, e.g., the progestin, estrogen, androgen, and glucocorticoid receptors, bind to heat-shock proteins and FKBP52, a calcium-regulated immunosuppressant, to form functional complexes (Peattie et al., 1992; Sanchez et al., 1990). DNA binding proteins and general transcription factors bind to the TATA box, a consensus sequence in a gene's promoter region that specifies the position of transcription initiation, forming a functional transcription complex (Chalut et al., 1995). Proteins can interact with multiple molecules simultaneously. For example, Nedd4, an ubiquitin-protein ligase, can interact with multiple proteins and lipids through its lipid binding domain and multiple protein binding domains (Jolliffe et al., 2000).

[099] Proteins utilize a large number of motifs to bind other molecules. Binding protein-related sequences can possess or interact with the cold-shock DNA-binding (CSD) domain, a conserved domain of about 70 amino acids that helps the cell survive in temperatures below optimum growth temperature by inducing the synthesis of proteins that negatively regulate transcription, translation, and recombination, resulting in suppressed cell proliferation (http://pfam.wustl.edu/cgi-bin/getdesc?name=CSD). Proteins induced by exposure to cold include DNA-binding proteins, and cold inducible RNA binding proteins, which have RNA binding domains at or near their N-termini (Nishiyama et al., 1997). For example, contrin, a testis-specific DNA/RNA binding protein with a cold shock domain also has a large number of phosphorylation sites, each of which can mediate intermolecular interactions (Tekur et al., 1999). Contrin is involved in transcription of testis-specific genes; its inactivation could provide a reversible male contraceptive.

[0100] Binding protein-related sequences can possess or interact with the ARID/BRIGHT DNA binding (ARID) domain, which is an approximately 100 amino acid sequence involved in a wide range of DNA interactions, including, but not limited to, interaction with AT-rich regions (http://pfam.wustl.edu/cgi-bin/getdesc?

name=ARID). ARID-encoding genes are involved in a variety of biological processes, including regulation of cell growth, development, cell lineage gene regulation, cell cycle control, and tissue-specific gene expression.

[0101] Binding protein-related sequences can also possess or interact with nucleosomal binding domains to facilitate binding within the nucleosome, a nuclear structure comprised of chromosomal DNA and proteins. For example, the HMG14 and HMG17 (HMG14_17) domain is present in some nucleosome proteins, most commonly, in proteins HMG14 and HMG17, members of a family designated as high mobility group proteins, which form components of chromatin, and bind to nucleosomal DNA, regulating the interaction of the DNA with histone proteins (http://pfam.wustl. edu/cgi-bin/getdesc? name=HMG14_17).

[0102] Binding protein-related sequences can also possess or interact with conserved motifs that recognize RNA, and allow the protein to bind RNA (http://pfam. wustl.edu/cgi-bin/textsearch?terms=rna+ binding&search_what= all&sections=DE&sections =CC&size=100). These motifs include the RNA recognition (rrm) domain, also known as a RRM, RBD, or RNP domain (http://pfam. wustl.edu/cgi-bin/getdesc?name=rrm). Numerous RNA binding proteins possess the rrm domain, including heterogeneous nuclear ribonucleoproteins (hnRNP) proteins, which are implicated in the regulation of alternative splicing, and LA proteins, which are among the main autoantigens in systemic lupus erythematosus (SLE).

[0103] Binding protein-related sequences can also possess or interact with conserved motifs that mediate their binding to ions, e.g., calcium. Calcium-binding proteins such as calmodulin, the calcineurins, and their homologues and related proteins are widely used to regulate cellular processes (http://pfam.wustl.edu/cgi-bin/textsearch?terms=calcium +binding& search_what=all&sections=DE&sections= CC&size=100). Ion-binding proteins include phosphoproteins that bind to other molecules in an manner dependent on their phosphorylation state, and can regulate many types of molecules and processes, including those that utilize complex signaling cascades (Pang et al., 2001; Pang et al., 2002; Lin et al., 1999). Ion-binding protein-related sequences can possess or interact with the EF hand (efhand) domain, a calcium-binding domain that comprises a loop of twelve amino acids that coordinates a calcium ion in a pentagonal bipyramidal configuration and is flanked on both sides by a twelve amino acid alpha-helical domain (http://pfam.wustl. edu/cgi-bin/getdesc? name=efhand).

Breakpoint-Related Sequences

[0104] A breakpoint is the location on a chromosome where a gene is disrupted, and one segment of the gene is severed from the other. Chromosomal breaks that disrupt coding or regulatory sequences can result in gene mutation. Chromosomal breaks can also serve as molecular landmarks, e.g., a break can be detected on Southern blots as the loss of an expected band and the appearance of two novel bands. Examples of breakpoint-related sequences include the sequences that generate the Philadelphia chromosome translocation, the sequences that generate the chromosome translocation (t(1;7)(q42;p15)), which is implicated in Wilms' tumor, and the sequences that generate the chromosomal translocation t(18;21)(q22.1q21.3), which is implicated in Down syndrome.

[0105] Breakpoints commonly occur in discrete regions of the chromosome. Breakage at these regions can lead to a recognized disease phenotype. One way of generating such a phenotype is by chromosomal translocation, i.e., chromosomes mutate by exchanging parts. When a segment from one chromosome is exchanged with a segment from another nonhomologous chromosome, two mutated chromosomes are simultaneously generated (Griffiths, et al., 1999). The Philadelphia chromosome, a mutation sometimes associated with chronic myelogenous leukemia (CML), is an example. It results from the translocation of a discrete segment of chromosome 22 into a discrete region of chromosome 9. Patients with the Philadelphia chromosome mutation generally have a better prognosis than CML patients with other characteristics.

[0106] Acquired clonal chromosomal abnormalities are found in the malignant cells of most patients with leukemia, lymphoma, and solid tumors. Some of these abnormalities are the result of consistent chromosomal rearrangements. For example, in a preponderant number of chronic myelogenous leukemia cases, breakpoints at chromosome band 22q11 occur within a breakpoint cluster region of 5-6 kb (Weinstein et al., 1988).

[0107] Chromosome rearrangements affecting band 3q21 are associated with a particularly poor prognosis in myeloid leukemia or myelodysplasia. These breakpoints cluster in a breakpoint cluster region of approximately 30 kb, located centromeric and downstream of the ribophorin I (RPN-I) gene (Weiser, 2002). The apoptotic gene *bcl-2*, was isolated as a breakpoint rearrangement in human follicular

lymphomas and was shown to act as an oncogene that promoted cell survival rather than cell proliferation.

[0108] Some proteins can act as leukemia or lymphoma-specific antigens for major histocompatibility complex-restricted T cell cytotoxicity. These include the breakpoint cluster region (bcr)-abl, and other fusion oncoproteins. Genetically engineered chimeric and humanized antibodies have demonstrated activity against overt lymphomas and leukemias. Radioimmunotherapy has produced significant therapeutic responses with minimal radiation exposure to normal tissues (Jurcic et al., 2000).

[0109] Breakpoint-related sequences can possess or interact with RhoGAP domains, also known as the breakpoint cluster region-homology domain, and mediates signal transduction by small G proteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=RhoGAP). Breakpoint-related sequences can also possess or interact with RhoGEF domains, which comprise approximately 200 amino acid residues that encode a guanine nucleotide exchange factor (http://pfam.wustl.edu/cgi-bin/getdesc?name=RhoGEF). Breakpoint-related sequences can also possess or interact with Plectin/S10 (S10_plectin) domains, which are found at the N-terminus of some isoforms of plectin and ribosomal S10 protein (http://pfam.wustl.edu/cgi-bin/getdesc?name=S10_plectin).

Carrier or Transport-Related Sequences

[0110] A membrane transport protein is an integral transmembrane protein that aids one or more molecules across a cell membrane. Most, if not all, types of molecules are transported across membranes, including proteins, ions, and fatty acids (Schaffer and Lodish, 1994). Even molecules such as water and urea, which can diffuse across pure phospholipid bilayers, are frequently accelerated by transport proteins. Transporters clear cells of toxins, and confer drug resistance on tumor lines (Ramalho-Santos et al., 2002). The rate of transport varies considerably among membrane transport proteins. Membrane transport proteins function in the plasma membrane and in intracellular organellar membranes, including the nuclear, mitochondrial, lysosomal, and vesicular membranes. For example, transportin, also known as karyopherin beta2, imports nuclear mRNA binding proteins from the cytoplasm across the nuclear membrane, into the nucleus (Bonifaci et al., 1997).

[0111] Membrane transport proteins can have either a broad or a narrow range of specificity for the transported substance. In mammalian cells, nucleoside

transport across membranes is mediated by broad specificity transporters. Nucleoside transport plays a role in such diverse cellular functions as nucleotide synthesis, neurotransmission, and platelet aggregation. Nucleoside transporters carry chemotherapeutic nucleosides, and are a target of interest in chemotherapeutic and cardiac drug design (Griffiths et al., 1997; Ku et al., 1990).

[0112] Carriers are another class of membrane transport proteins; they bind to a solute and transport it across the membrane by undergoing a series of conformational changes. In contrast to channel proteins, transporters bind only one, or a few, substrate molecules at a time; after binding substrate molecules, they undergo a conformational change such that the bound substrate molecules, and only those molecules, are transported across the membrane. Carriers transport a wide variety of molecules, including fatty acids across the plasma membrane (Schaffer and Lodish, 1994); purines, pyrimidines, and components of nucleosides across the nuclear membrane, and adenine nucleotides across the inner mitochondrial membrane (Battini et al., 1997).

[0113] Membrane transport-related sequences can possess or interact with vacuolar (H^+)-ATPase C, D, G, and H subunit (V-ATPase) domains, which are membrane-attached sequences that generate an acidic environment (http://pfam.wustl.edu/cgi-bin/getdesc? name=V-ATPase_C). Membrane transport-related sequences can also possess or interact with nucleoside transporter (nucleoside_tran) domains, which are found in proteins that transport nucleosides across the plasma membrane, and are employed to synthesize nucleotides via the salvage pathways in cells that lack their own de novo synthesis pathways (http://pfam.wustl.edu/cgi-bin/getdesc?name=Nucleoside_tran). Membrane transport-related sequences can also possess or interact with ATP synthase F/14-kDa subunit (ATP-synt-F) domains, which correspond to a 14-kDa subunit in the peripheral catalytic part of vacuolar ATPases (http://pfam.wustl.edu/cgi-bin/getdesc? name=ATP-synt_F). Membrane transport-related sequences can also possess or interact with mitochondrial carrier protein (mito_carr) domains, which are involved in energy transfer across the inner mitochondrial membrane (http://pfam.wustl.edu/cgi-bin/getdesc?name=mito_carr). Membrane transport-related sequences can also possess or interact with an AMP-binding enzyme (AMP-binding) domain, which is a domain rich in serine, threonine, and glycine, and is characterized by a conserved

proline-lysine-glycine triplet sequence (http://pfam.wustl.edu/cgi-bin/getdesc?name=AMP-binding).

[0114] Membrane transport proteins, such as those expressed in cancer cells, are useful as targets for therapeutic intervention, for example, in the screening for small molecule inhibitors. Inhibition of membrane transport, as indicated above, may make cancer cells more susceptible to chemotherapy, for example.

Channel-Related Sequences

[0115] Channel proteins transport water or specific types of ions down their concentration or electrical potential gradients. They form a protein-lined passageway across the membrane through which multiple water molecules or ions move at a very rapid rate, e.g., up to 10^8 per second. The plasma membrane, for example, contains potassium-specific channel proteins that generate the cell's resting electric potential across the plasma membrane. Examples of channel-related sequences include the sodium hydrogen exchanger, sodium potassium ATPase, and the cystic fibrosis transmembrane regulator.

[0116] Members of this subset of membrane transport proteins have wide-ranging functions in both normal physiology and in pathology. For example, the transport system that mediates the transmembrane exchange of sodium for hydrogen across the plasma membrane plays a physiological role in the regulation of intracellular pH, the control of cell growth and proliferation, stimulus-response coupling, metabolic responses to hormones, the regulation of cell volume, and the transepithelial absorption and secretion of several ions. The sodium-hydrogen exchanger also plays a role in cancer and in tissue and organ hypertrophy (Mahnensmith and Aronson, 1985).

[0117] Channel-related sequences can possess or interact with sodium/hydrogen exchanger (Na_H_Exchanger) domains, which exchange sodium for hydrogen across a membrane in an electroneutral manner (http://pfam.wustl.edu/cgi-bin/getdesc? name=Na_H_Exchanger). Channel-related sequences can also possess or interact with neurotransmitter-gated ion-channel ligand binding (Neur_chan_LBD) domains, which form the extracellular domains of some ion channels (http://pfam.wustl.edu/cgi-bin/getdesc?name=Neur_chan_LBD). Channel-related sequences can also possess or interact with UBX domains, which are present in ubiquitin-regulatory proteins (http://pfam.wustl.edu/ cgi-bin/getdesc?name=UBX).

Checkpoint-Related Sequences

[0118] The cell division cycle is the fundamental means by which living things are propagated. Fundamental to successful propagation is the faithful replication of DNA; a cell cycle control system exists to coordinate the cycle as a whole. The control system is regulated by brakes that can stop the cycle at specific checkpoints. Thus, the checkpoints arrest the cycle upon the occurrence of undesirable events, such as DNA damage, replication stress, or mitotic spindle disruption. For example, DNA lesions and disrupted replication forks are recognized by the DNA damage checkpoint and replication checkpoint, respectively. Checkpoints can also, for example, initiate protein kinase-based signal transduction cascades to activate downstream effectors that elicit cell cycle arrest, DNA repair, or apoptosis. These actions prevent the conversion of aberrant DNA structures into inheritable mutations and minimize the survival of cells with unrepairable damage (Qin and Li, 2003).

[0119] Dysregulation of the cell-cycle is a hallmark of tumor cells. Defective checkpoint function results in genetic modifications that contribute to tumorigenesis. Checkpoint function can be abrogated by many different mechanisms (Bast, et al., 2000). For example, cyclin-dependent kinases that normally are activated at a checkpoint can be inactivated or activated in an abnormal manner. Alternatively, the normal activities of the cyclin-dependent kinase inhibitors, phosphatases, or other regulatory molecules of the cell cycle can be altered. Tumor suppressors are among the classes of molecules that can effect cell cycle dysregulation. The abrogation of checkpoint function can alter the sensitivity of tumor cells to chemotherapeutics (Stewart et al, 2003).

[0120] Checkpoint-related sequences can possess or interact with phosphoribosylaminoimidazole-succinocarboxamide synthase (SAICAR_synt) domains, which function in *de novo* purine synthesis (http://pfam.wustl.edu/cgi-bin/getdesc?name =SAICAR_synt). Checkpoint-related sequences can also possess or interact with WD40 domains, which comprise a domain of approximately 40 amino acids, which are sometimes present in tandem repeats (http://pfam.wustl.edu/cgi-bin/getdesc?name=WD40). Checkpoint-related sequences can also possess or interact with cyclin, C-terminal (cyclin_C) domains, which regulate cyclin dependent kinases (http://pfam.wustl.edu/cgi-bin/getdesc? name=cyclin_C).

[0121] Thus, checkpoint related proteins, e.g., kinases, phosphatases, etc., are useful as targets for therapeutic intervention, such as in screening for small molecule drugs for the treatment of cancer, immune disorders, and inflammation.

Complex-Related Sequences

[0122] Complexes are molecular entities comprised of two or more components. Molecular complexes within cells form functional units that carry out cellular operations. For example, complexes at the cell membrane perform structural and regulatory tasks, including regulating membrane traffic and maintaining organelle integrity. Complexes at the cytoskeleton perform static and dynamic roles with respect to cell shape, intracellular transport, and communication with the extracellular matrix. Complexes in the nucleus transcribe and regulate genes, and complexes at sites of protein synthesis translate and regulate proteins. Complexes can reside intracellularly and/or extracellularly, e.g., in the extracellular matrix. Examples of complex-related sequences include cytoskeletal and filamentous proteins, ADP-ribosylation factor (ARF) proteins, and protein synthesis initiation factors (Amor et al., 1994).

[0123] Complex-related sequences can possess or interact with ADP-ribosylation factor family (arf) domains, which are GTP-binding domains involved in protein trafficking (http://pfam.wustl.edu/cgi-bin/getdesc?name=arf). Complex-related sequences can also possess or interact with eukaryotic initiation factor domains, e.g., the eukaryotic initiation factor 4E (IF4E) domain, which recognizes and binds mRNA during protein synthesis (http://pfam.wustl.edu/cgi-bin/getdesc?name=IF4E). Complex-related sequences can also possess or interact with intermediate filament (filament) protein domains, which form filamentous structures typically 8 to 14 nm wide, and form components of the cytoskeleton and nuclear envelope, e.g., neurofilaments, cytokeratins, lamins, vimentin, and desmin (http://pfam.wustl.edu/cgi-bin/getdesc?name=filament).

Cytokine-Related Sequences

[0124] A cytokine is an extracellular signaling protein or peptide that acts as a local mediator in communication among cells. Cytokines regulate proliferation and differentiation, for example, they mediate differentiation of cells in the hematopoeitic lineage. Examples of cytokines include interleukins, interferons, and colony stimulating factors of the hematopoeitic system. Some cytokines, e.g., interferons and interleukins, can be induced by viral activity, and possess antiviral activity (Sheppard

et al., 2003). Cytokine-related sequences may enable the expression of a cytokine, for example, as a cytokine transcription factor (Kao et al., 1994). They can also be part of a cytokine effector pathway, for example, as an intracellular effector of cytokine-related cytoskeletal changes in response to events in the extracellular matrix (Hirsh et al., 2001; Joberty et al., 1999).

[0125] Cytokine-related sequences can possess or interact with interferon-induced transmembrane protein (CD225) domains, which are associated with interferon-induced cell growth suppression (http://pfam.wustl.edu/ cgi-bin/getdesc?name=CD225). Cytokine-related sequences can also possess or interact with SelR (SelR) domains, which bind both selenium and zinc, and/or methionine sulfoxide reductase enzymatic domains (http://pfam. wustl.edu/cgi-bin/getdesc?name=SelR). Cytokine-related sequences can also possess or interact with reverse transcriptase (rvt) domains, which are involved in RNA-directed DNA polymerase activity, an enzymatic activity that uses an RNA template to produce DNA for integration into a host genome (http://pfam.wustl.edu/cgi-bin/getdesc?name=rvt). Cytokine-related sequences can also possess or interact with L1 transposable element domains (Transposase_22), which are described above.

[0126] Cytokines, thus, are useful as therapeutic proteins for the treatment of disorders such as cancer, immune disorders, and inflammation.

Dehydrogenase-Related Sequences

[0127] Dehydrogenases are enzymes that catalyze the removal of hydrogen atoms in the absence of oxygen. They contribute to a wide range of enzymatic reactions, including those involved in amino acid degradation, amino acid synthesis, the citric acid cycle, fatty acid oxidation, fatty acid synthesis, glycolysis, the pentose phosphate pathway, photosynthesis, pyruvate oxidation, and oxidative phosphorylation (Walker et al., 1992). Examples of dehydrogenases include steroid dehydrogenases, NADH dehydrogenases, and glyceraldehyde-3-phosphate dehydrogenase.

[0128] Dehydrogenase-related sequences can possess or interact with glyceraldehyde 3-phosphate dehydrogenase, NAD binding (GPDH) domains, which play a role in glycolysis and gluconeogenesis by reversibly catalyzing the oxidation and phosphorylation of D-glyceraldehyde-3-phosphate to 1,3-diphospho-glycerate (http://pfam.wustl.edu/cgi-bin/getdesc?name=gpdh). Dehydrogenase-related sequences can also possess or interact with 3-hydroxyacyl-CoA dehydrogenase, NAD

binding (3HCDH_N) domains, which catalyze the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA in fatty acid metabolism (http://pfam.wustl.edu/cgi-bin/getdesc?name=3HCDH_N).

Disease-Related Sequences

Amyotrophic Lateral Sclerosis

[0129] Amyotrophic Lateral Sclerosis (Lou Gehrig's Disease) is a neurodegenerative disease that affects the motor neurons. The disease displays multiple clinical variants and can affect motor neurons throughout the nervous system, e.g., the spinal cord and brainstem. One clinical variant, the autosomal recessive form of juvenile amyotrophic lateral sclerosis, has been mapped to the human chromosome 2q33-q34 region (Hadano et al., 2001). A protein family characterized by the HAP1 N-terminal conserved region (HAP1_N) domain possesses a N-terminal conserved region from hypothetical protein products of ALS2CR3 genes found in the 2q33-2q34 region of chromosome 2 (http://pfam.wustl.edu/cgi-bin/getdesc?name= HAP1_N).

Gaucher's Disease

[0130] Gaucher's Disease is a genetic disease characterized by a deficiency of enzymes responsible for the breakdown and recycling of glycolipids, i.e., lipids with carbohydrate moieties, e.g., glucosylceramide; and sphingolipids, lipids with sphingosine moieties, e.g., sphingomyelin. Normally, the glycolipids and sphingolipids in the membranes of senescent cells are metabolized by a multi-step process that includes the activities of acid beta-glucosidases and saposins. When these activities are absent, or present in reduced amounts, glucosylceramide and sphingolipids accumulate, and produce the Gaucher's disease phenotype. The disease displays multiple clinical variants, and can manifest with central nervous system pathology, enlargement of organs, e.g., liver and spleen, and an increase in the level of the cytokine transforming growth factor beta (Zhao and Grabowski, 2002; Perez Calvo et al., 2000; Cormand et al., 1997). The variability in clinical presentation is consistent with the large number of different mutations observed in the acid beta-glucosidase and saposin genes.

[0131] Acid beta-glucosidases are enzymes that metabolize glycolipids. Saposins are small proteins that are described in more detail below. Mammalian saposins are synthesized as a single precursor molecule (prosaposin) with saposin-A (SAPA) and saposin-B (SapB_1; SapB_2) domains; prosaposin becomes an active

saposin following a proteolytic activation reaction (http://pfam.wustl.edu/cgi-bin/getdesc?name=SAPA; http://pfam.wustl. edu/cgi-bin/getdesc?name=SapB_1; http://pfam.wustl.edu/cgi-bin/getdesc?name=SapB_1).

Huntington Disease

[0132] Huntington Disease is a progressive neurodegenerative genetic disorder characterized by dementia, psychiatric symptoms, and a choriform movement disorder. It is caused by an increased number of repeats of the codon CAG, which encodes the amino acid glutamine, in a gene located at the 4p16.3 region of chromosome 4, which codes for a protein called huntingtin. The polyglutamine tracts expressed by the mutant form of the gene selectively ablate striatal and cortical neurons, (Ho et al., 2001).

[0133] The Huntington Disease gene is widely expressed, but exerts tissue-specific effects on neurons (Lin et al., 1993). The gene expresses multiple distinct transcripts, and differential polyadenylation of the gene leads to the expression of transcripts of different sizes (Lin et al., 1993). There is a relative increase in the abundance of one transcript in the human brain, which has been hypothesized to account for the tissue-specific effects of the disease (Lin et al., 1993). The HAP1_N protein domain, described above, binds to the gene product, huntingtin, in a polyglutamine repeat-length-dependent manner (http://pfam.wustl.edu/cgi-bin/getdesc?name=HAP1_N). This domain is also found in several huntingtin-associated protein 1 (HAP1) homologues.

Multiple Sclerosis (MS)

[0134] Multiple sclerosis (MS) is a disease characterized by demyelination, i.e., the loss of the myelin coating, of nerve axons. Its clinical course varies among patients; these variations fall into two broad categories, a relapsing/remitting course, and a chronic progressive course. MS has a complex etiology; it has an autoimmune component, is influenced by genetics, and sometimes involves infectious agents. MS results from an abnormal immune response to one or more antigens present in the myelin sheaths that cover the nerve axons of genetically susceptible individuals, which may be preceded by exposure to a causal infectious agent (Oksenberg et al., 1999).

[0135] The genetic susceptibility to MS is determined by MS susceptibility genes, most of which demonstrate only a small to moderate effect on susceptibility, e.g., the major histocompatibility complex at chromosome 6p21 (Oksenberg et al.,

1999). An etiological infectious agent has been isolated from the plasma and cerebrospinal fluid of patients with multiple sclerosis (Perron et al., 1997). This agent is a retroviral oncovirus, known as multiple sclerosis-associated retrovirus (MSRV), also called LM7, and is found in association with virions produced by the cultured cells of MS patients (Perron et al., 1997). MSRV proteins possess protein domains characteristic of retroviral proteins. These include the Gag P30 core shell protein (Gag_p30) domain, which is involved in viral assembly (http://pfam.wustl.edu/cgi-bin/getdesc?name=Gag_p30) and the reverse transcriptase (rvt) domain, which was described above.

Obesity

[0136] Although single-gene mutations have been shown to cause obesity in animal models, the most common forms of human obesity arise from the interactions of multiple genes, environmental factors, and behavior. Several genes have been shown to affect body weight regulation in humans and other animals. These include the ob, lep, CPE, ASIP, LEP, TUB, UPC, POMC, CCKAR, TNFA, and PPAR-γ genes (Comuzzie et al., 1998). Genetic regulation of body weight can be effected through diverse mechanisms. For example, the TUB gene family regulates body weight by encoding proteins that are phosphorylated in response to insulin, mediate insulin signaling, and are associated with a maturity onset obesity associated with insulin resistance (Ikeda et al., 2002). CCKAR genes regulate body weight in a different manner; they regulate the hormone cholecystokinin, which produces a feeling of satiety following food intake (Ritter et al., 1994).

[0137] Some genes that regulate body weight possess the WH1 domain, which is described above. Genes that regulate body weight can also possess or interact with the sprouty (sprouty) domain. This domain is found in sprouty proteins, which inhibit the Ras/mitogen-activated protein kinase cascade, a pathway initiated by receptor tyrosine kinases and involved in development (http://pfam.wustl.edu/cgi-bin/getdesc?name=Sprouty). Genes that regulate body weight can also possess or interact with a Tub (Tub) domain, which is found in Tubby, a mouse gene in which an autosomal recessive mutation resulting from a splicing defect causes maturity-onset obesity, insulin resistance and sensory deficits (http://pfam.wustl.edu/cgi-bin/getdesc?name=Tub).

Oncogene

[0138] An oncogene is any one of a large number of genes that can help make a cell cancerous. Typically, an oncogene is a mutant form of a normal gene, and is often a gene involved in the control of cell growth, division, or differentiation. Cells in higher organisms normally grow, divide, differentiate, and die under the regulation of other cells. Cancer cells proliferate, in part, because they are able to divide without input from other cells, as the result of accumulated mutations. Oncogenes include, but are not limited to, genes encoding GTP binding proteins, e.g., *ras*; growth factors, e.g., platelet-derived growth factor; growth factor receptors, e.g., platelet-derived growth factor receptor; kinases, e.g., *src*; nuclear proteins, e.g., *myc*; and tumor suppressors, e.g., retinoblastoma proteins.

[0139] The products of oncogenes are frequently proteins involved in cell signaling, e.g., kinases, GTP-binding proteins, and receptors. For example, many human cancers have a mutation in a ras gene (Alberts et al., 1994). The ras proteins belong to a large superfamily of monomeric GTPases, and relay signals from receptor tyrosine kinases to the nucleus, stimulating cell proliferation or differentiation. Ras proteins function as switches, cycling between an active state in which GTP is bound, and an inactive state, in which GDP is bound. A ras gene mutation can result in the translation of a protein that fails to hydrolyze its bound GTP, and persists abnormally in its active state, transmitting an intracellular signal for cell proliferation or differentiation even in the presence of regulatory non-proliferation and non-differentiation signals. Oncogene-related proteins can possess one of many ras protein domains (http://pfam.wustl.edu/cgi-bin/textsearch?terms=ras&search_what=all&sections=DE &sections=CC&size=100), including the sub-families Ras, Rab, Rac, Ral, Ran, Rap, and Ypt1. Oncogene-related proteins can also possess a Gtr1/RagA G-protein conserved region (gtr1_RagA) domain, which is found in some G-proteins of the Ras family, e.g., the RagA/B human homologues of the ras GTP binding protein Gtr1 (http://pfam.wustl.edu/cgi-bin/getdesc?name=Gtr1_RagA). Oncogene-related sequences can also possess or interact with an ATPase domain associated with diverse cellular activities; proteins with the AAA ('A'TPases 'A'ssociated with diverse cellular 'A'ctivities) domain can perform chaperone-like functions that assist in assembling, operating, or disassembling protein complexes. The domain includes a conserved region of approximately 220 amino acids that contains an ATP-binding site which can act as an ATP-dependent protein clamp to

hold a protein in place (http://pfam.wustl.edu/cgi-bin/getdesc?name=AAA). Some oncogene-related sequences can also possess or interact with a C2 domain of approximately 116 amino-acid residues, which can be involved in calcium-dependent phospholipid binding and inositol-1,3,4,5-tetraphosphate binding, and is found, e.g., in some isozymes of protein kinase C (http://pfam.wustl.edu/cgi-bin/getdesc?name=C2). C2 domains are typically located between C1 domains (which bind phorbol esters and diacylglycerol) and protein kinase catalytic domains. Regions with homology to the C2 domain are present in many proteins, e.g., synaptotagmin.

Parkinson's Disease

[0140] Parkinson's disease is a neurological disorder that affects movement control. Complex interactions among groups of nerve cells in the central nervous system coordinate to control movement. One such group of neurons is located in the substantia nigra of the midbrain; these neurons release the neurotransmitter dopamine, which allows an organism to fine-tune its movements. In Parkinson's disease, neurons of the substantia nigra progressively degenerate, leaving the patient with clinical symptoms that may include resting tremor, muscular rigidity, a slowness of spontaneous movement, and poor balance and motor coordination (Seigel et al., 1999).

[0141] Parkinson's disease has multiple causes, including both genes and the environment. It also has multiple presentations, including juvenile-onset (before age 45) and adult onset (after age 45), and can be transmitted through either autosomal dominant or autosomal recessive mechanisms. In keeping with the diversity of etiologies, presentation, and genetic mechanisms, there are a large and diverse number of genes and gene products involved in the pathogenesis of Parkinson's disease. For example, the PARK2 gene, which encodes the protein parkin, is mutant in autosomal recessive juvenile parkinsonism. PARK2 is a ubiquitin protein ligase that is a component in the pathway that attaches ubiquitin to specific proteins, designating them for degradation (Fishman, and Oyler, 2002).

[0142] Parkinson's disease-related sequences can possess or interact with synuclein domains, which are expressed on the cytoplasmic regions of proteins found predominantly in neurons (http://pfam.wustl.edu/cgi-bin/getdesc?name=Synuclein). Alpha-synuclein, which possesses a synuclein domain, is mutated in several families with autosomal dominant Parkinson's disease. Gamma-synuclein, which also

possesses a synuclein domain, is overexpressed in breast and ovarian cancers (Lavedan, 1998).

Retinitis Pigmentosa

[0143] Retinitis pigmentosa is a group of inherited retinopathies characterized by early stage loss of night vision, followed by loss of peripheral vision. Defects in any structural or functional proteins associated with the rod photoreceptor neurons of the retina, which are the cells that transduce light into a neuronal action potential, can lead to the disease (Seigel et al., 1999).

[0144] GTPase regulators have been implicated in the pathology of retinitis pigmentosa. GTPase regulators are proteins that determine whether a GTP binding protein exists in a GTP-bound or GDP-bound state (Zhao et al., 2003); they are described in more detail below. GTPase regulators have a broad spectrum of intracellular functions, including intracellular vesicular transport. These proteins localize to a specific region of rod photoreceptor cells, in a narrow cilium that connects the cell body, where protein synthesis and basic metabolism takes place, with the rod outer segment, where light is transduced to an action potential of the optic nerve (Zhao et al., 2003). Proteins necessary for the light transduction process are made in the cell body and must be transported to the outer segment via vesicular transport mechanisms. Mutant GTPase regulators, which regulate vesicular transport, play a role in the pathogenesis of retinitis pigmentosa (Roepman et al., 2000). Retinitis pigmentosa-related sequences can possess or interact with a Tctex-1 domain, which is comprised of a dynein light chain, and can bind to the cytoplasmic tail of rhodopsins, which are light-sensing proteins present in retinal rod cells (http://pfam.wustl. edu/cgi-bin/getdesc?name=Tctex-1). Mutations in this domain that are responsible for retinitis pigmentosa inhibit this binding.

Alzheimer's Disease

[0145] Alzheimer's disease is a neurodegenerative dementing illness. It is a genetically complex disease with multiple forms, including familial and sporadic forms, and early onset and late-onset forms. Mutations in at least four genes are known to cause Alzheimer's disease, and there is evidence for additional Alzheimer's loci (McKusick, 2003). One form of Alzheimer's disease is caused by mutations in the amyloid precursor gene, another form is associated with the apolipoprotein E4 allele, a third form is caused by a mutant presenilin-1 gene that encodes a seven-transmembrane domain protein, and a fourth form is caused by a mutant gene

encoding a similar seven -transmembrane domain protein, presenilin-2 (McKusick, 2003).

[0146] Consistent with its multiple etiologies, multiple clinical presentations, and multiple genetic loci, Alzheimer disease has a complex pathology. One facet of the pathology of Alzheimer's disease is the formation of amyloid plaques from amyloid precursor protein (Clark and Karlawish, 2003). Amyloid precursor protein can be processed *in vitro* by several different proteases such as secretases and caspases to yield peptide fragments, suggesting that these proteases may play a role in the formation of pathogenic amyloid plaques *in vivo* (Suh and Checler, 2002). Presenilins have been identified as likely candidates for the proteases that cleave amyloid precursor protein to pathogenic peptide fragments *in vivo* (Selkoe, 2001). Another facet of Alzheimer's disease pathology is an inflammatory component mediated by microglial cells, the brain's primary immunoeffector cells (Tan et al., 1999). Microglial cells are attracted to and activated by amyloid deposits; they release inflammatory mediators that promote the aggregation of the deposits into plaques, and also directly induce or promote neurodegeneration (Hoozemans et al., 2002). Therefore, current treatment strategies include anti-inflammatory and immunotherapeutic approaches, including vaccines (Weiner and Selkoe, 2002).

[0147] Alzheimer's disease-related sequences can possess or interact with trypsin domains, which demonstrate a wide range of peptide degrading activities, including exopeptidase, endopeptidase, oligopeptidase and omega-peptidase activities (http://pfam. wustl.edu/cgi-bin/getdesc?name=trypsin). Alzheimer's disease-related sequences can also possess or interact with low-density lipoprotein receptor (ldl_rece) domains, which are characterized by seven successive cysteine-rich repeats of about 40 amino acids at the N-terminal region, and which are also present in receptors for low density lipoprotein (LDL), the major cholesterol-carrying lipoprotein of plasma (http://pfam.wustl.edu/cgi-bin/textsearch?terms=ldl_rece +&search_what=all& sections =DE&sections=CC&size=100). Alzheimer's disease-related sequences can also possess or interact with a PT repeat (pt_a) domain, which includes the tetrapeptide XPTX, or a similar, conserved, sequence.

Williams-Beuren Syndrome

[0148] Williams-Beuren syndrome is a complex genetic developmental disorder with multisystemic manifestations, and variability in its presentation. In 90-95% of the cases reported, a gene deletion occurs at the 7q11.23 location on the long

arm of chromosome 7; in the remaining cases, a variety of other chromosomal deletions and translocations have been observed (Wang et al., 1999). The most severe cases are characterized by cardiac anomalies, including aortic stenosis, mental retardation, growth deficiency, a characteristic facial appearance, dental malformation, and infantile hypercalcemia (Lashkari et al., 1999).

[0149] The underlying molecular basis for the syndrome is the absence of the proteins encoded by the genes of the affected region of the chromosome. A missing elastin gene, with resulting extracellular matrix anomalies, is a consistent finding. Other genes that are present in and near the commonly deleted region of chromosome 7, and thus are likely to contribute to pathogenesis, are (1) a gene encoding a regulator of chromosome condensation-like G-exchanging factor, which is a factor that exchanges nucleotides for small GTP-binding proteins, (2) an N-acetylgalactosaminyltransferase, (3) a DNAJ-like chaperone, (4) NOL1/NOP2/sun domain-containing proteins, including a novel protein designated WBSCR20, which is expressed in skeletal muscle, and is similar to a 120 kilodalton proliferation-associated nucleolar antigen, (5) a methyltransferase designated WBSCR22, and (6) other proteins with no known homologies (Merla et al., 2002; Doll and Grzeschik, 2001). Williams-Beuren-related sequences can possess or interact with a GTF2I-like repeat (GTF2I) domain, which is a DNA binding domain commonly deleted in Williams-Beuren syndrome, (http://pfam.wustl.edu/cgi-bin/getdesc?name=GTF2I).

Rheumatic Diseases

[0150] Rheumatic diseases are inflammatory conditions that can have autoimmune, infective, or traumatic origins. They include arthritis, systemic lupus erythematosus, scleroderma, and Sjogren's syndrome. Arthritis refers to any inflammation of a joint. Systemic lupus erythematosus is an autoimmune disease in which patients produce antibodies to their own tissues, resulting in an inflammatory process that can damage organs. Scleroderma can present as systemic scleroderma, a chronic, progressive disease that is characterized by hardening and stiffening of the skin and damage to internal organs, e.g., heart, lungs, kidneys and esophagus. Sjogren's syndrome is a progressive immunological disorder characterized by inflammation and the subsequent destruction of exocrine glands, e.g., salivary glands, sweat glands, and lacrimal (tear) glands.

[0151] The serum of patients with scleroderma and Sjogren's syndrome have antibodies directed against a protein that is a normal component of the Golgi

apparatus (Seelig et al., 1994), an intracellular organelle composed of a stack of flattened cisternae with associated transport vesicles. The Golgi apparatus sorts proteins and sends them to their correct intracellular destination. This antigenic protein is a "golgin," one of a class of molecules characterized by an integral membrane domain and a large cytoplasmic region. Golgins organize the Golgi's structure, and influence protein sorting (Gillingham et al., 2002). Golgins function in a variety of ways, including cross-bridging Golgi cisternae to one another (Linstedt and Hauri, 1993) and tethering Golgi transport vesicles to the cisternal membranes (Shorter et al., 2002). Rheumatic disease-associated sequences can possess or interact with golgin-97, RanBP2alpha, Imh1p, and p230/golgin (GRIP) domains, which are found in many large coiled-coil proteins, are sufficient for targeting to the Golgi, and have a conserved tyrosine residue (http://pfam.wustl.edu/cgi-bin/getdesc?name=GRIP).

Disintegrin-Related Sequences

[0152] Disintegrins are proteins that interfere with the function of integrins. Disintegrins are generally proteins of about 70 amino acid residues that contain multiple disulfide bonds, bind with high affinity to a subset of integrins, and interfere with integrin binding to physiological ligands. Examples of disintegrin-related sequences include snake venoms and related proteins, cysteine-rich metalloproteinases and related non-enzymatic sequences, e.g., those expressed in the male reproductive tract, and membrane-anchored metalloproteinases with diverse functions, e.g., the shedding of cell-surface proteins such as cytokines and cytokine receptors, and the conferring of asthma susceptibility (Van Eerdewegh et al., 2002; Perry et al., 1995).

[0153] Disintegrin-related sequences can possess or interact with disintegrin domains, which contain an Arg-Gly-Asp sequence, a sequence commonly found in adhesion proteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=disintegrin). Proteins that comprise both disintegrin and metalloproteinase peptidase domains include ADAM proteins. Disintegrin-related sequences can also possess or interact with reprolysin family propeptide (Pep_M12B_propep) domains, which are domains that include the propeptide sequence of members of the peptidase family M12B, and contain a sequence motif similar to a sequence found in matrixin proteins (http://pfam.wustl.edu/ cgi-bin/getdesc?name=Pep_M12B_ propep).

Factor-Related Sequences

[0154] A factor is any molecule that contributes to a bodily process. Factors can function in specific biochemical reactions and cellular functions. There are many categories of factors, and factors are involved in many, if not all, physiological and pathological processes. Some exemplary factors are described in the following paragraphs; they are not exhaustive of the category.

[0155] Transcription factors are factors that initiate or regulate transcription in eukaryotes. They include gene regulatory proteins, which turn specific sets of genes on or off, and general transcription factors, which assemble at the promoter region to enable and regulate transcription of many genes. They also include transcription elongation factors, which are proteins required for the addition of amino acids to growing polypeptide chains on ribosomes (Alberts et al., 1994). Transcription factors interact with a wide variety of molecules, including DNA binding proteins, polymerases, regulatory molecules such as kinases, and specific regions of DNA, e.g., promoters, and enhancers (Alberts et al., 1994; Vallejo et al., 1993).

[0156] Translation factors, including translation initiation factors and release factors, are involved in initiating and regulating the rate of protein synthesis. They also interact with many molecules, including ribosomal proteins, mRNA, and molecules that regulate the incorporation of amino acids into protein, such as kinases and GTP (Price et al., 1993; Alberts, 1994).

[0157] Export factors are involved in the export of molecules, e.g., RNA, from the nucleus (Stutz et al., 2000). Folding factors are involved in the process of folding proteins into their functional three dimensional shapes, and are also involved in receptor function (Gao et al., 1994). Factors such as activators and coactivators interact with nuclear receptors to modulate cellular processes, e.g., transcription (Mahajan et al., 2002).

[0158] ADP-ribosylation factors are involved in the addition of an ADP-ribose group donated from nicotinamide adenine dinucleotide (NAD) to specific amino acid residues in heterotrimeric G-proteins. They are involved in, for example, normal cellular processes, such as vesicular transport, and also in the pathologic states induced by cholera, pertussis, and botulinum toxins (Alberts et al., 1994; Amor et al., 1994). Guanine nucleotide exchange factors bind to small G-proteins, such as Ras, and displace GDP in favor of GTP. They act as effectors or modulators of small G-proteins (Ehrhardt et al., 2001; Janeway et al., 2001; Shao and Andres, 2000).

[0159] Factor-related sequences can possess or interact with ADP-ribosylation factor family (arf) domains, which are GTP-binding domains involved in protein trafficking (http://pfam.wustl.edu/cgi-bin/getdesc?name=arf). Factor-related sequences can also possess or interact with elongation factor Tu GTP binding (GTP_EFTU) domains, which are elongation factors that promote the GTP-dependent binding of aminoacyl tRNA to ribosomes during protein biosynthesis, and catalyze the translocation of the newly synthesised protein chain (http://pfam.wustl.edu/cgi-bin/getdesc?name=GTP_EFTU). Factor-related sequences can also possess or interact with 4F5 protein family (4F5) domains, which comprise ubiquitously expressed short proteins rich in aspartate, glutamate, lysine and arginine (http://pfam.wustl.edu/cgi-bin/getdesc?name=4F5). Factor-related sequences can also possess or interact with eukaryotic initiation factors, e.g., eukaryotic initiation factor 4E (IF4E), which recognizes and binds mRNA during an early step of protein synthesis (http://pfam.wustl.edu/cgi-bin/getdesc?name=IF4E).

Germ Cell Specific Protein-Related Sequences

[0160] Germ cells, also called gametes, are cells that contribute to a new generation of organisms by giving rise to either an egg or a sperm. They are haploid cells specialized for sexual fusion. Proteins that are specific to germ cells can be found at one or more developmental stages of gametes.

[0161] Germ cell-related sequences include germ cell genes and their gene products, their regulators and effectors, genes and gene products affected in disorders associated with germ cells, and antibodies that specifically recognize or modulate germ cell-related sequences. Examples of germ cell-related sequences include the germ cell-specific Y-box binding protein and contrin. Germ cell specific protein-related sequences possess or interact with the cold-shock DNA-binding (CSD) domain, which is described above.

Growth Factor-Related Sequences

[0162] A growth factor is an extracellular polypeptide signaling molecule that stimulates a cell to grow or proliferate. Many types of growth factors exist, including protein hormones and steroid hormones. Some growth factors have a broad specificity, and some have a narrow specificity. Examples of growth factors with broad specificity include platelet-derived growth factor, epidermal growth factor, insulin like growth factor I, transforming growth factor β, and fibroblast growth

factor, which act on many classes of cells. Examples of growth factors with narrow specificity include erythropoeitin, which induces proliferation of precursors of red blood cells, interleukin-2, which stimulates proliferation of activated T-lymphocytes, interleukin-3, which stimulates proliferation and survival of various types of blood cell precursors, and nerve growth factor, which promotes the survival and the outgrowth of nerve processes from specific classes of neurons.

[0163] Most growth factors have other actions in addition to inducing cell growth or proliferation, e.g., they may influence survival, differentiation, migration, or other cellular functions. Growth factors can have complex effects on their targets, e.g., they may act on some cells to stimulate cell division, and on others to inhibit it. They may stimulate growth at one concentration, and inhibit it an another. Growth factors are also involved in tumorogenesis.

[0164] Growth factor related sequences include sequences associated with the process of stimulating cell growth or proliferation by a growth factor. For example, they include intracellular effectors of growth, such as components of intracellular pathways that respond to growth factors (Kothapalli et al., 1997; Wax et al., 1994), sequences that bind directly or indirectly to growth factors (Van den Berghe et al., 2000), and sequences affected as a result of growth factor action.

[0165] Growth factor-related sequences can possess or interact with a transforming growth factor beta like (TGF-beta) domain, which is a multifunctional peptide sequence that controls proliferation, differentiation and other functions in many cell types (http://pfam. wustl.edu/cgi-bin/getdesc?name=TGF-beta). Growth factor-related sequences can also possess or interact with a fibroblast growth factor (FGF) domain, which is found in a family of proteins involved in growth and differentiation (http://pfam.wustl.edu/cgi-bin/getdesc? name=FGF).

GTPase-Related Sequences

[0166] GTPases are enzymes that catalyze GTP hydrolysis, and comprise a large family of proteins with a similar globular GTP binding domain. When GTP is bound to a GTPase, it is hydrolyzed to GDP, and the domain undergoes a conformational change that inactivates the protein. GTPases are regulated by GTPase regulators, proteins that determine whether a GTP binding protein exists in a GTP-bound or GDP-bound state (Zhao et al., 2003). GTPase regulators include GTPase activating proteins, which bind the GTPase and induce it to hydrolyze its bound GTP to GDP; the GTPase remains in an inactive, GDP-bound state until it

encounters a guanine nucleotide releasing protein, which binds to the GTPase and causes the release of the nucleotide. GTPases have a broad spectrum of intracellular functions, including intracellular vesicular transport. Examples of GTPase-related sequences include ras, GTPase-activating proteins, and guanine nucleotide releasing proteins.

[0167] GTPase-related sequences can possess or interact with GTPase activator protein for Ras-like GTPase (RasGAP) domains, which are protein domains of about 250 residues that accelerate the GTPase activity of ras (http://pfam.wustl.edu/cgi-bin/getdesc?name=RasGAP). GTPase-related sequences can also possess or interact with putative GTPase activating protein for ARF (ArfGap) domains, which are protein domains with a zinc finger involved in intermolecular associations (http://pfam.wustl.edu/cgi-bin/getdesc?name=ArfGap). GTPase-related sequences can also possess or interact with ankyrin repeat domains (ank), which are tandemly repeated modules of about 33 amino acids found in a variety of functionally diverse proteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=ank). GTPase-related sequences can also possess or interact with pleckstrin homology (PH) domains, which are protein domains of about 100 residues involved in intracellular signaling, or as components of the cytoskeleton (http://pfam.wustl.edu/cgi-bin/getdesc?name=PH).

Heat-Shock Protein-Related Sequences

[0168] Heat-shock proteins, also referred to as stress-response proteins, are proteins that are synthesized in response to an elevated temperature or other cell stressor, and help the cell withstand environmental insults. A cell stressor can induce a battery of genes that encode gene products that protect the cell from the result of the insult, e.g., proteins that stabilize and repair partially denatured cell proteins. Some heat-shock proteins, e.g., chaperones, are present at high levels in unstressed cells, and further induced by stress. Chaperones assist other proteins in attaining their proper secondary and tertiary structures. For example, members of the tubulin-specific chaperone A family possess tubulin-specific chaperone A (TBCA) domains that fold tubulin polypeptides into their functional configuration (http://pfam.wustl.edu/cgi-bin/getdesc?name=TBCA).

[0169] Heat and other stressors further induce the synthesis of a family of 90-kDa heat-shock proteins that are already abundant in unstressed cells (Pepin et al., 2001;Lees-Miller et al., 1989; Rebbe et al., 1987). Members of this family possess a hsp 90 protein (HSP90) domain that interacts with tubulin, actin, tyrosine kinase

oncogene products of retroviruses, eIF2alpha kinase, and steroid hormone receptors (Lees-Miller and Anderson, 1989). This domain includes a highly-conserved N-terminal region, separated from a conserved, acidic C-terminal region by a highly-acidic, flexible linker region (http://pfam. wustl.edu/cgi-bin/getdesc?name=HSP90).

[0170] Another family of heat-shock proteins, the hsp70 proteins, have an average molecular weight of 70 kDa; some members of this family are only expressed under conditions of stress, while some are present in cells under normal conditions. Hsp70 proteins reside in different cellular compartments, e.g., the nucleus, cytosol, mitochondria, and endoplasmic reticulum. Hsp70 proteins, e.g., Hsc73, can be differentially expressed at different stages of development (Soulier et al., 1996). Hsp70 proteins, e.g., the chaperone hsp70-like dnaK protein, can associate with proteins that possess a DnaJ domain, which comprises an N-terminal conserved domain of about 70 amino acids, a glycine-rich region of about 30 amino acids, a central domain containing four repeats of a CXXCXGXG motif, and a C-terminal region of 120 to 170 amino acids (http://pfam.wustl.edu/cgi-bin/getdesc?name=DnaJ). Proteins with DnaJ domains can be postranslationally modified by farnesylation (Andres et al., 1997).

Helicase-Related Sequences

[0171] Helicases are enzymes that use energy from the hydrolysis of ATP to unwind the DNA helix at the replication fork, allowing the single stands to be copied. Proteins with DNA helicase activity play roles in DNA replication, repair, and recombination. Disorders associated with helicases include Xeroderma pigmentosum, Cockayne syndrome, diffuse collagen disease, alpha-thalassemia, Bloom syndrome, Werner syndrome, and Rothmund-Thomson syndrome (Miyajima, 2002). Examples of helicases include RNA helicases, RECQL4, and minichromosome maintenance helicase.

[0172] Helicase-related sequences can possess or interact with helicase associated (HA) domains, which are protein domains comprising alpha helices that may bind to nucleic acids (http://pfam.wustl.edu/cgi-bin/getdesc?name=HA). Helicase-related sequences can also possess or interact with helicase conserved C-terminal (helicase_C) domains, which are protein domains that are found in a subset of helicases designated the DEAD/H helicases (http://pfam.wustl.edu/ cgi-bin/getdesc?name=helicase_C).

Hydrolase-Related Sequences

[0173] Hydrolases are enzymes that catalyze the hydrolysis of a variety of bonds, such as esters, glycosides, and peptides. Hydrolases split a molecule into fragments by adding water; the water's hydrogen atom is incorporated into one fragment, and the hydroxyl group is incorporated into another. Hydrolases are involved in a wide range of physiological and pathological processes, including proteolysis, phosphatase activity, and sugar metabolism. Examples of hydrolases include protein hydrolases, lipid hydrolases, nucleic acid hydrolases, and small molecule, e.g., coenzyme A, hydrolases (Hawes et al., 1996).

[0174] Hydrolase-related sequences can possess or interact with alpha/beta hydrolase fold (abhydrolase) domains, which are catalytic domains found in a wide range of hydrolytic enzymes of different phylogenetic origins and catalytic functions (http://pfam.wustl.edu/cgi-bin/getdesc?name=abhydrolase). Hydrolase-related sequences can also possess or interact with dUTPase domains, which are proteins domains that hydrolyze dUTP to dUMP and pyrophosphate.

Immune Cell-Related Sequences

[0175] An immune cell is a cell involved in, or associated with, the immune system. Immune cells include cells in the myeloid and lymphocytic arms of the immune response, as well as their precursors. Immune cells also include cells at all stages in the differentiation pathways that produce cells associated with the immune system. These cells can reside, either permanently or temporarily, in the spleen, lymph nodes or mucosal-associated lymphoid tissues (MALT). Immune cell-related sequences are involved in all functions of the immune response, e.g., antibody production and cell-mediated immunity, and can function at any point in time, ranging from the embryonic formation of the immune system, through the time of an immune challenge, to many decades later, e.g., when a B-cell memory response is invoked (Janeway, 2001).

[0176] Immune-cell related sequences of differentiating immune cells include pre-B cells that do not produce immunoglobulin light chain, but express a transcript homologous to immunoglobulin lambda light-chain genes, the expression of which is limited to pre-B cells and select other cells that have no surface immunoglobulin (Hollis et al., 1989). Immune-cell related sequences of activated immune cells include a B-cell-restricted transcription factor expressed by activated B

cells; its expression pattern suggests it has a role in regulating B-cell differentiation (Massari et al., 1998).

[0177] Examination of the expression of immune-cell related sequences can detect and diagnose immunoregulatory abnormalities. For example, genes that encode proteins which mediate the combinatorial process that combines a finite number of component genes into the very broad range of antigen-specific immunoglobulin and T-cell binding proteins, are expressed at higher levels in patients with systemic lupus erythematosis (SLE) than in healthy subjects (Girschick et al., 2002).

[0178] Immune cell-related sequences can possess or interact with a CUB domain, which is an extracellular domain of approximately 110 amino acids, and is present in functionally diverse, including developmentally regulated, proteins (http://pfam.wustl.edu/ cgi-bin/getdesc?name=CUB). Immune cell-related sequences can also possess or interact with a CD-20 domain, which has four transmembrane regions, both extracellular and cytoplasmic extensions, and is found, inter alia, in a high affinity IgE receptor (http://pfam.wustl.edu/cgi-bin/getdesc?name=CD20). Immune cell-related sequences can also possess or interact with an interferon-induced transmembrane protein (CD225) domain, which is found in a family of proteins that includes the human leukocyte antigen CD225, an interferon-inducible transmembrane protein associated with interferon-induced cell growth suppression (http://pfam.wustl.edu/cgi-bin/getdesc?name=CD225). Immune cell-related sequences can also possess or interact with sushi domains, also known as complement control protein (CCP) modules, or short consensus repeats (SCR). These domains are found in a wide variety of complement and adhesion proteins, including proteins responsible for the antigenicity of blood group antigens on the external face of the red blood cell membrane (http://pfam.wustl.edu/cgi-bin/getdesc?name=sushi). Immune cell-related sequences can also possess or interact with SH2 domains and rvt domains; both are described above.

Integrase-Related Sequences

[0179] Integrases are enzymes that form proviruses by inserting a linear double-stranded DNA copy of a retroviral genome into host cell DNA. Examples of integrases include HIV integrase, PhiC31 integrase, and Sip.

[0180] Integrase-related sequences can possess or interact with an integrase zinc binding domain (Integrase_Zn) domain, which is a zinc binding protein

domain placed near the N-terminus (http://pfam.wustl.edu/cgi-bin/getdesc?name=Integrase_Zn). Integrase-related sequences can also possess or interact with an integrase core (rve) domain, which is a protein domain that forms the central catalytic core of the integrase (http://pfam.wustl.edu/ cgi-bin/getdesc?name=rve). This domain acts as an endonuclease to cleave the nucleotide and catalyzes the transfer of the viral DNA strand to the integration site of the host DNA. Integrase-related sequences also possess or interact with an integrase DNA binding (integrase) domain, which is a DNA-binding protein domain near the C-terminus (http://pfam.wustl.edu/cgi-bin/getdesc?name=integrase). Integrase-related sequences also possess or interact reverse transcriptase (rvt) domains, which are described above. Integrase-related sequences also possess or interact with a RNase H domain, which is a protein domain that hydrolyzes the RNA portion of RNA/DNA hybrids (http://pfam.wustl.edu/cgi-bin/getdesc?name=rnaseH).

Integrin-Related Sequences

[0181] Integrins are transmembrane proteins that mediate cell to cell as well as cell to matrix adhesion, and provide a means of communication between the interior of a cell and the extracellular matrix. The extracellular portion of integrins binds to components of the extracellular matrix, e.g., collagen, fibronectin and laminin. The intracellular portion of integrins interacts with the cell cytoskeleton, e.g., actin filaments near the cell surface. Integrins transmit information about the extracellular environment across the plasma membrane to the cytoskeleton, where it is available to intracellular signaling mechanisms (Alberts et al., 1994). Structurally, integrins consist of heterodimers of an alpha and a beta subunit. Each subunit has a large N-terminal extracellular domain followed by a transmembrane domain and a short C-terminal cytoplasmic region. The pairing of certain alpha subunits with certain beta-subunits determines ligand specificity, localization and function. The extracellular binding domains of integrins often bind their ligands with low affinity; simultaneous, weak, binding with multiple matrix molecules provides the cell with a means to sense its complex, changing, extracellular environment without becoming glued to it. Examples of integrin-related sequences include integrin alpha and beta subunits, collagens, and integrin-linked kinase (Zhang et al., 2002).

[0182] Integrin-related sequences can possess or interact with von Willebrand factor type A (vwa) domains, which are protein domains that participate in diverse biological functions, e.g., cell adhesion, migration, homing, pattern

formation, and signal transduction (http://pfam.wustl. edu/cgi-bin/getdesc?name=vwa). Integrin-related sequences can also possess or interact with FG-GAP repeat (FG-GAP) domains, which are protein domains present in the vicinity of ligand binding domains at the N-terminus of integrin alpha subunits (http://pfam.wustl.edu/cgi-bin/getdesc?name=FG-GAP).

Interacting Protein-Related Sequences

[0183] An "interacting protein" is a protein that interacts with another molecule. Interacting proteins are involved in every aspect of cellular function. Interacting proteins have been characterized in all known locations in the cell, and include all, or most types of, proteins. Interacting proteins in the nucleus regulate such diverse functions as apoptosis, transcription, homologous recombination, and DNA repair. Nuclear fibroblast growth factor-2 interacting factor interacts with fibroblast growth factor 2 to prevent apoptosis (Van den Berghe et al., 2000). Grap2 cyclin-D interacting protein (GCIP) a nuclear cell-cycle protein, inhibits select transcriptional events, and reduces the leve l of phosphorylation of nuclear retinoblastoma protein (Chang et al., 2000). Pir 51, a human homologue of Rec A, a bacterial enzyme that mediates genetic recombination, interacts with the enzyme rad51 to regulate homologous recombination and DNA repair in mammalian cells (Kovalenko et al., 1997). Hepatitis B virus X-associated protein (HBXAP), a protein demonstrated to play a role in the development of hepatocelluar carcinoma, interacts with the hepatitis B virus regulatory gene product HBx to increase viral transcription (Shamay et al., 2002).

[0184] Interacting protein-related proteins can utilize many protein domain motifs for interaction. They can possess or interact with domains that mediate interaction with DNA, RNA, ions, or other proteins. For example, PDZ domains, which are also known as DHR or GLGF domains, target signaling molecules to membranes and mediate the assembly of functional membrane domains (Fanning and Anderson, 1999). Interacting protein-related proteins can also possess or interact with rrm domains, which are described above.

Isomerase-Related Sequences

[0185] Isomerases are enzymes that convert molecules into their positional isomers, i.e., into molecules with the same chemical formula but a different stereochemical arrangement of atoms. Isomerases act on a wide variety of molecules, including sugars, amino acids, and nucleic acids. They are involved in a wide range

of physiological and pathological functions, including those involving metabolic and synthetic pathways.

[0186] Isomerase-related sequences include isomerase genes and gene products, their substrates, products, activators, inhibitors, effectors, and cofactors, regulatory molecules that modulate their function, genes and gene products affected in disorders associated with isomerases and antibodies that specifically recognize or modulate isomerase-related sequences. Examples of isomerase-related sequences include triosephosphate isomerases, peptidyl-prolyl isomerases, glucose phosphate isomerases, disulfide isomerases, ketosteroid isomerases, and ribosyltransferase-isomerases (Brown et al., 1985).

[0187] Isomerase-related sequences can possess or interact with triosephosphate isomerase (TIM) domains, which are protein domains that catalyze the reversible interconversion of glyceraldehyde 3-phosphate and dihydroxyacetone phosphate (http://pfam.wustl.edu/cgi-bin/getdesc?name=TIM). Isomerase-related sequences can also possess or interact with cyclophilin type peptidyl-prolyl cis-trans isomerase (pro_isomerase) domains, which accelerate protein folding by catalyzing the cis-trans isomerization of peptide bonds (http://pfam.wustl.edu/ cgibin/getdesc?name=pro_ isomerase).

Mucin-Related Sequences

[0188] The term mucin refers to both an albumin-like substance that is present in mucus, and to transmembrane proteins that can typically be produced in both soluble and transmembrane forms. Soluble mucins comprise mucus gels that protect epithelial cells in the airways, digestive tract, and other organs, and are found in body fluids, such as milk, tears, and saliva. In their transmembrane forms, mucins provide a steric barrier to protect the apical surface of epithelial cells. Transmembrane mucins are also involved in pathogenesis; for example, they mediate viral entry into cells, promulgate the inflammatory response, and are involved in the regulation of abnormal cell proliferation (Jeffery and Zhu, 2002; Tsuda et al., 1993). Examples of mucins include MUC2 mucin, mucin carcinoembryonic antigen, and Muc3 membrane bound intestinal mucin.

[0189] Mucin -related sequences can possess or interact with mucin-like glycoprotein (tryp_mucin) domains, which are domains that are involved in the interaction of parasites with host cells (http://pfam.wustl.edu/cgi-bin/getdesc?name=Tryp_mucin). Mucin-related sequences can also possess or

interact with multi-glycosylated core protein (MGC-24) domains, which are protein domains of sialomucins that are expressed in many normal and cancerous tissues (http://pfam.wustl.edu/cgi-bin/getdesc?name=MGC-24).

Other Polypeptide-Related Sequences

[0190] In addition to the sequences described above, the sequences of the invention include nucleotide and amino acid sequences, some with known function, and some with unknown function, that fall into a broad array of categories. These sequences are listed below in SEQ ID NOS.: 210 - 418, as "Other Polypeptides with Known Function," and "Other Polypeptides," respectively.

[0191] Polypeptide-related sequences of the invention can possess or interact with groucho/TLE N-terminal Q-rich (TLE_N) domains, which are protein domains found in co-repressor proteins, and are involved in oligomerization (http://pfam.wustl.edu/cgi-bin/getdesc?name=TLE_N). Polypeptide-related sequences of the invention can also possess or interact with uncharacterized protein family 0160 (UPF0160) domains, which are protein domains found in proteins that include multiple metal-binding residues, and in some cases act as a phosphodiesterase (http://pfam.wustl.edu/cgi-bin/getdesc?name=UPF0160). Polypeptide-related sequences of the invention can also possess or interact with SNF7 domains, which are protein domains involved in protein sorting and transport from the endosome to the lysosome or vacuole of eucaryotic cells (http://pfam.wustl.edu/cgi-bin/getdesc?name=SNF7). Polypeptide-related sequences of the invention can also possess or interact with NifU-like N-terminal (NifU_N) domains, which are protein domains involved in nitrogen fixation, and other functions (http://pfam.wustl.edu/cgi-bin/getdesc? name=NifU_N). Polypeptide-related sequences of the invention can also possess or interact with tRNA synthetases class II (D, K, and N) (tRNA-synt_2) domains, which are protein domains that activate the amino acids asparagines, aspartic acid, and lysine, and transfer them to specific tRNA molecules (http://pfam.wustl.edu/cgi-bin/getdesc?name=tRNA-synt_2).

[0192] Polypeptide-related sequences of the invention can also possess or interact with dynein heavy chain (dynein_heavy) domains, which are protein domains that correspond to the C-terminal region of the dynein heavy chain (http://pfam.wustl.edu/cgi-bin/getdesc?name=Dynein_heavy). Polypeptide-related sequences of the invention can also possess or interact with cyclin-dependent kinase regulatory subunit (CKS) domains, which are protein domains of approximately 79-

150 amino acid residues that are involved in regulating progression through the cell cycle (http://pfam.wustl.edu/cgi-bin/getdesc?name= CKS).

[0193] Polypeptide-related sequences of the invention can also possess or interact with nucleoside diphosphate linked to some other moiety X (NUDIX) domains, which are protein domains that are involved in removing oxidatively damaged nucleotides (http://pfam.wustl.edu/cgi-bin/getdesc?name=NUDIX). Polypeptide-related sequences of the invention can also possess or interact with T-complex protein/cpn60 chaperonin (cpn60_TCP1) domains, which are protein domains involved in protein folding and oligomerization (http://pfam.wustl.edu/cgi-bin/getdesc?name=cpn60_TCP1). Polypeptide-related sequences of the invention can also possess or interact with F-actin capping protein, beta subunit (F_actin_cap_B) domains, which are protein domains of approximately 280 amino acids that are involved in capping actin, i.e., blocking the exchange of actin monomers (http://pfam.wustl.edu/cgi-bin/getdesc?name=F_actin_cap_B).

[0194] Polypeptide-related sequences of the invention can also possess or interact with G-protein alpha subunit (G-alpha) domains, which are protein domains that bind guanyl nucleotides, and function as a GTPase (http://pfam.wustl.edu/cgi-bin/getdesc? name=G-alpha). Polypeptide-related sequences of the invention can also possess or interact with Kruppel-associated box (KRAB) domains, which are protein domains involved in protein-protein interactions, and present in some zinc finger proteins (http://pfam.wustl.edu/ cgi-bin/getdesc?name=KRAB). Polypeptide-related sequences of the invention can also possess or interact with metallopeptidase family M24 (Peptidase_M24) domains, which are protein domains that are found in some metalloproteases, including proline dipeptidase, and methionine aminopeptidase (http://pfam.wustl.edu/cgi-bin/getdesc?name=Peptidase_M24). Polypeptide-related sequences of the invention can also possess or interact with thioredoxin (thiored) domains, which are protein domains involved in oxidation/reduction reactions by reversibly oxidizing disulfide bonds (http://pfam.wustl.edu/cgi-bin/getdesc? name=thiored).

[0195] Polypeptide-related sequences of the invention can also possess or interact with TUDOR domains, which are protein domains involved in the formation of primordial germ cells, and for normal abdominal segmentation (http://pfam.wustl.edu/cgi-bin/getdesc?name =TUDOR). Polypeptide-related sequences of the invention can also possess or interact with SIT4 phosphatase-

associated protein (SAPS) domains, which are protein domains that are involved in cyclin transcription (http://pfam.wustl.edu/cgi-bin/getdesc?name=SAPS). Polypeptide-related sequences of the invention can also possess or interact with ankyrin repeat (ank) domains, which are protein domains of approximately 33 amino acids, and are sometimes found in tandemly repeated modules (http://pfam.wustl.edu/ cgi-bin/getdesc? name=ank). Polypeptide-related sequences of the invention can also possess or interact with nicotinamide N-methyltransferase/phenylethanolamine N-methyltransferase/ thioether S-methyltransferase (NNMT_PNMT_TEMT) domains, which are protein domains that are found in proteins that use S-adenosyl-L-methionine as the methyl donor (http://pfam.wustl.edu/cgi-bin/getdesc?name= NNMT_PNMT_TEMT). Polypeptide-related sequences of the invention can also possess or interact with C1q domains, which are protein domains involved in activating the serum complement system (http://pfam.wustl.edu/cgi-bin/getdesc? name=C1q). Polypeptide-related sequences of the invention can also possess or interact with collagen triple helix repeat (Collagen) domains, which are protein domains that typically form extracellular connective tissue (http://pfam.wustl.edu/cgi-bin/getdesc? name=Collagen).

[0196] Polypeptide-related sequences of the invention can also possess or interact with the hyaluronan/mRNA binding family (HABP4_PAI-RBP1) domain, which is a protein domain that can bind to the glucosaminoglycan hyaluronan, and to RNA (http://pfam.wustl.edu/cgi-bin/getdesc?name=HABP4_PAI-RBP1). Polypeptide-related sequences of the invention can also possess or interact with eucaryotic aspartyl protease (asp) domains, which are protein domains that cleave peptide bonds; proteins with this domain include pepsins, cathepsins, and rennin (http://pfam.wustl.edu/cgi-bin/getdesc?name=asp). Polypeptide-related sequences of the invention can also possess or interact with trypsin domains, which are protein domains that function as serine proteases (http://pfam.wustl.edu/ cgi-bin/getdesc? name=trypsin). Polypeptide-related sequences of the invention can also possess or interact with Kunitz/Bovine pancreatic trypsin inhibitor (Kunitz_BPTI) domains, which are protein domains that is found in serine protease inhibitors (http://pfam.wustl.edu/cgi-bin/getdesc?name=Kunitz_BPTI). Polypeptide-related sequences of the invention can also possess or interact with proliferating cell nuclear antigen, N-terminal (PCNA) domains, which are protein domains that are found on non-histone

acidic nuclear proteins, and play a role in controlling DNA replication (http://pfam.wustl.edu/cgi-bin/getdesc?name=PCNA).

Oxygenase-Related Sequences

[0197] Oxygenases are enzymes that catalyze the incorporation of molecular oxygen into organic substances. Dioxygenases, also known as oxygen transferases, catalyze the introduction of both atoms of molecular oxygen, and typically contain iron. Monooxygenases, also known as mixed function oxygenases, introduce one oxygen atom; the other is reduced to water. Examples of oxygenase-related sequences include cytochrome oxygenases, heme oxygenases, cyclooxygenases, lipoxygenases, and peptide-aspartate beta-dioxygenase.

[0198] Oxygenase-related sequences can possess or interact with alkyl hydroperoxide reductase/thiol specific antioxidant (AhpC-TSA) domains, which are responsible for providing a defense against sulfur-containing radicals; proteins that possess this domain include allergens, e.g., asp f 3, mal f 2, and mal f 3 (http://pfam.wustl.edu/cgi-bin/getdesc?name=AhpC-TSA). Oxygenase-related sequences can also possess or interact with monooxygenase domains, which are protein domains that utilize flavin adenine dinucleotide (FAD) (http://pfam.wustl.edu/cgi-bin/getdesc?name=Monooxygenase). Oxygenase-related sequences can also possess or interact with dioxygenase domains, which are protein domains that catalyze the incorporation of both atoms of molecular oxygen into substrates (http://pfam.wustl.edu/cgi-bin/getdesc?name= Dioxygenase).

Peroxidase-Related Sequences

[0199] Peroxidases are enzymes that catalyze the reduction of hydrogen peroxide. Peroxidases are generally located within peroxisomes, which are intracellular organelles that metabolize fatty acids and toxic compounds. Disorders associated with peroxidase-related sequences include X-linked adrenoleukodystrophy. Examples of peroxidase-related sequences include glutathione peroxidases, thiol peroxidases, catalases, horseradish peroxidases, anionic peroxidases, and thyroid peroxidases.

[0200] Peroxidase-related sequences can possess or interact with alkyl hydroperoxide reductase/thiol specific antioxidant (AhpC-TSA) domains, which are protein domains that can reduce organic hydroperoxides (http://pfam.wustl.edu/cgi-bin/getdesc? name=AhpC-TSA).

Phospholipase-Related Sequences

[0201] Phospholipases are enzymes that act on phospholipids. They characteristically generate products that are active in signal transduction pathways. For example, phospholipase C hydrolyzes phosphatidylinositol bisphosphate (PIP_2) to generate the two intracellular mediators, inositol trisphosphate (IP_3) and diacylglycerol. IP_3 releases Ca^{2+} from stores in the endoplasmic reticulum, increasing the cytosolic Ca^{2+} concentration. Diacylglycerol remains in the plasma membrane and activates protein kinase C.

[0202] Phospholipase activity is involved in the synthesis of eicosanoids, inflammatory mediators that include prostaglandins, prostacyclins, thromboxanes, and leukotrienes. Corticosteroid hormones, such as cortisone, for example, inhibit phospholipase activity in the first step of the eicosanoid synthesis pathway. Corticosteroid hormones are widely used clinically to treat noninfectious inflammatory diseases, such as some forms of arthritis (Ribardo et al., 2002).

[0203] Phospholipids play a pivotal role in the modulation of intestinal inflammation. The mucosal surface of the digestive tract functions as a regulatory barrier between the gastrointestinal lumen and the underlying mucosal immune system. Phospholipids help preserve the mucosa following various forms of injury or physiological damage to the lumen, thus preventing invasion of harmful luminal factors into the host, which subsequently may lead to inflammation, or a pathological immune response, both promoting and inhibiting gastrointestinal inflammation and immunity (Sturm and Dignass, 2002).

[0204] Phospholipase-related sequences can possess or interact with lysophospholipase catalytic (PLA2_B) domains, which catalyze the release of fatty cids from lysophospholipids (http://pfam.wustl.edu/cgi-bin/getdesc?name=PLA2_B). Phospholipase-related sequences can also possess or interact with phospholipase/carboxylesterase (abhydrolase_2) domains, which have broad substrate specificity (http://pfam.wustl.edu/cgi-bin/getdesc?name=abhydrolase_2). Phospholipase-related sequences can also possess or interact with GDSL-like lipase/acylhydrolase (Lipase_GDSL) domains, which are present in lipolytic enzymes with serine in the active site (http://pfam.wustl.edu/cgi-bin/getdesc?name= Lipase_GDSL).

Prosaposin-Related Sequences

[0205] Saposins are small lysosomal proteins that activate lysosomal lipid-degrading enzymes, including enzymes that metabolize sphingosine. They typically isolate lipids from their membrane surroundings, and increase their accessibility to degradative enzymes. Mammalian saposins are synthesized as a single precursor molecule, prosaposin, which becomes an active saposin following proteolytic activation. Examples of prosaposin-related sequences include saposin A, saposin B, and saposin C. Disorders associated with prosaposin-related sequences include neurodegenerative diseases similar to similar to Tay-Sachs and Sandhoff diseases, e.g., Gaucher's disease, which is described above.

[0206] Prosaposin-related sequences can possess or interact with saposin-A (SAPA) domains, saposin B1 (SapB_1) domains, and saposin B2 (SapB_2) domains, which are described above.

Proteasome-Related Sequences

[0207] Proteasomes are intracellular complexes that degrade proteins. Proteasomes recognize proteins that have been marked for destruction by the addition of an ubiquitin molecule, unfold these ubiquitinated proteins, cleave them into small peptides of 6-12 amino acids, and release them into the cytosol (Mitch and Goldberg, 1996). Examples of proteasome-related sequences include 26S proteasome subunits, 26S proteasome regulatory chains, and ubiquitin.

[0208] Proteasome-related sequences can possess or interact with proteasome/cyclosome repeat (PC_rep) domains, which are protein domains that are present in regulatory subunits of the proteasome (http://pfam.wustl.edu/cgi-bin/getdesc?name= PC_rep). Proteasome-related sequences can also possess or interact with Mov34/MPN/PAD-1 family (Mov34) domains, which are protein domains found at the N-terminus of regulatory subunits of the proteasome (http://pfam.wustl.edu/cgi-bin/getdesc?name=Mov34).

Reductase-Related Sequences

[0209] Reductases are enzymes that catalyze reduction reactions, i.e., reactions in which hydrogen is combined with a molecule, or reactions in which oxygen is removed from a molecule. Examples of reductases include dehydrogenase reductases, oxidoreductases, quinone reductases, CoA reductases, dihydrofolate reductases, tetrahydrofolate reductases, carbonyl reductases, nitrate reductases,

epoxide reductases, NADP(+) reductases, ribonucleotide reductases, and thioredoxin reductases (Loeffen et al., 1998).

[0210] Reductase-related sequences can possess or interact with short chain dehydrogenase (adh_short) domains, which are present in a wide variety of proteins (http://pfam.wustl.edu/cgi-bin/getdesc?name=adh_short). Reductase-related sequences can possess or interact with NADH-Ubiquinone oxidoreductase (complex I), chain 5 N-terminus (oxidored_q1_N) domains, which are protein domains that catalyze the transfer of electrons from NADH to ubiquinone in a reaction that can be associated with proton translocation across a membrane (http://pfam.wustl.edu/cgi-bin/getdesc?name=oxidored_q1_N).

Reverse Transcriptase-Related Sequences

[0211] Reverse transcriptases are enzymes that make double stranded DNA copies from single stranded nucleic acid template molecules. Typically, a reverse transcriptase is a DNA polymerase that can copy both RNA and DNA templates, and has an integral RNase H activity (Lim et al., 2002). The two enzymatic domains of reverse transcriptase reflect these two activities; the first is a DNA polymerase domain that can use either RNA or DNA as a template to synthesize either the minus-strand or the plus strand of DNA, and the second is an RNase H domain that degrades the RNA in RNA-DNA hybrids (Coffin, 1997; Wu and Gallo, 1975).

[0212] Reverse transcriptase plays a role in the replication of some viruses, e.g., retroviruses. It copies the retroviral RNA genome to produce a single minus strand of DNA, then catalyzes the synthesis of a complementary plus strand. Accordingly, reverse transcriptase is a therapeutic target for conditions that involve retroviruses, e.g., Aquired Immune Deficiency Syndrome (AIDS). A number of anti-retroviral drugs inhibit reverse transcriptase (Frank, 2002).

[0213] Reverse transcriptase is also a standard scientific research tool in the field of molecular biology. The reverse transcriptase polymerase chain reaction (RTPCR) amplifies specific DNA sequences rapidly, and *in vitro*. RTPCR can detect trace amounts of RNA and DNA, and is used in a wide range of applications, including forensics, the diagnosis of genetic diseases, determination of the prognosis of diagnosed diseases, and the detection of viral infection (Alberts, et al., 1994). For example, reverse transcriptase is used to diagnose cancer (Rowland, 2002), and to

provide prognostic information about the predicted survival of patients with prostate cancer (Kantoff et al., 2001).

[0214] An example of a reverse transcriptase is telomerase, a general tumor marker with a reverse transcriptase catalytic subunit (Kirkpatrick and Mokbel, 2001). Most human somatic cells do not express the telomerase reverse transcriptase gene; conversely, most cancer cells express this gene (Ducrest et al., 2002; Kyo et al., 2000). The human telomerase reverse transcriptase promoter has been placed in gene therapy vectors that specifically target telomerase-positive tumor cells, and spare nearby telomerase-negative cells (Pan and Koeneman, 1999). Human telomerase reverse transcriptase is also recognized as a tumor antigen that can be a target for immunotherapeutic approaches to cancer (Gordan and Vonderheide, 2002).

[0215] Reverse transcriptase-related sequences can possess or interact with rvt, transposase_22, WD40, and Exo_endo_phos domains, all of which are described above.

Ribosome-Related Sequences

[0216] A ribosome is a particle comprised of ribosomal proteins and ribosomal RNA that catalyzes protein synthesis from messenger RNA. Ribosomes are composed of two subunits, the large (L) subunit and the small (S) subunit. The typical mammalian ribosome comprises four RNA molecules and approximately eighty different proteins, which are highly conserved among prokaryotes and eukaryotes, and perform a variety of tasks related to protein synthesis . e.g., coordinating protein synthesis in a manner that maintains cell homeostasis (Yoshihama et al., 2002; Kenmochi et al., 1998).

[0217] Ribosomal proteins can perform functions independent of their involvement in protein synthesis. For example, they are involved in cell-cycle progression, e.g., as cell cycle checkpoints, and mediators of homologous recombination, embryogenesis, and skeletal development (Yoshihama et al., 2002; Chen and Ioannou, 1999). They also contribute to the regulation of cell growth, transformation, and death, and can induce apoptosis (Chen and Ioannou, 1999; Naora et al., 1999). Mutations in ribosomal proteins are associated with human diseases, including Down syndrome, Diamond-Blackfan anemia, Turner syndrome, and Noonan syndrome (Yoshihama et al., 2002).

[0218] Ribosomal proteins have been grouped into protein families on the basis of sequence similarities in functional domains. One family of ribosomal

proteins, the ribosomal protein L11, RNA binding (Ribosomal_L11) domain, is comprised of members that possess the L11 RNA binding domain; this family includes the ribosomal proteins L11 and L12, which are components of the large subunit. L11 is a protein of 140 to 165 amino-acids that binds to a 23S RNA molecule, the C-terminal region of which is buried within the ribosomal structure (http://pfam.wustl.edu/cgi-bin/getdesc?name=Ribosomal_L11). Another family of large ribosomal subunit proteins possess the ribosomal protein L13e (Ribosomal_L13e) domain, which is found in a wide range of vertebrates and in lower-order species (http://pfam.wustl.edu/cgi-bin/getdesc?name=Ribosomal_L13e), as is the ribosomal protein L44 (Ribosomal_L44) domain (http://pfam.wustl.edu/cgi-bin/getdesc?name= Ribosomal_L44).

[0219] Additional ribosomal protein families encompass small subunit proteins. The ribosomal protein S6e (Ribosomal_S6e) domain is present in a family of proteins which includes protein kinase substrates that/control cell growth and proliferation by selectively translating particular classes of mRNA (http://pfam.wustl.edu/cgi-bin/getdesc?name= Ribosomal_S6e). The ribosomal protein S8e (Ribosomal_S8e) domain is present in a family of proteins comprising approximately 220 amino acids in eukaryotes, and about 125 amino acids in archebacteria (http://pfam.wustl.edu/cgi-bin/getdesc?name=Ribosomal_S8e). The ribosomal protein S10p/S20e (Ribosomal_S10) domain is present in a family of proteins which includes the small ribosomal subunit S10 from prokaryotes and S20 from eukaryotes (http://pfam.wustl.edu/cgi-bin/getdesc?name= Ribosomal_S10). S10 is involved in binding transfer RNA to the ribosome, and also operates as a transcriptional elongation factor.

RNase-Related Sequences

[0220] RNases are enzymes that cleave RNA. RNases generally recognize their targets by tertiary structure, rather than by sequence; they include exonucleases, which remove the terminal base in an RNA sequence, and endonucleases, which can cleave non-terminal bases. Examples of RNases include RNase E, which is involved in the formation of 5S ribosomal RNA from pre-ribosomal RNA; RNase F, which cleaves both viral and host RNA in response to interferons, inhibiting protein synthesis; RNase H, which is specific for the RNA strand of an RNA-DNA hybrid; RNase P, which generates transfer RNA from

precursor transcripts; and RNase T, which removes the terminal AMP from nonaminoacylated tRNA (Coffin, et al., 1997).

[0221] RNase-related sequences can possess or interact with rvt, rve, RNase H, and gag_p30 domains, all of which are described above.

RNase H-Related Sequences

[0222] RNase H is a nuclease specific for the RNA strand of an RNA-DNA hybrid that cleaves phosphodiester bonds to produce molecules with 3'-OH and 5'-PO_4 ends. Multiple forms of RNase H are present in both prokaryotes and eukaryotes. RNase H may be part of larger polypeptides and its activity can be influenced by other regions of these polypeptides (Coffin, et al., 1997; Crouch 1990).

[0223] During retroviral replication, RNase H activity forms oligonucleotides that prime DNA synthesis. Therefore, the RNase H activity of reverse transcriptase is a target for therapeutic intervention. For example, small molecule inhibitors of retroviral RNase H function have shown promise in managing HIV infection (Klarman, et al., 2002).

[0224] Another therapeutic indication for RNase H is the regulation of cancer genes by targeting mRNA translation. Antisense deoxyoligonucleotides down-regulate mRNA expression by annealing to specific regions of an mRNA. Formation of the DNA:RNA heteroduplex then triggers mRNA cleavage by RNase H. Cleavage is rapidly followed by further degredation, irreversibly preventing translation of the target mRNA. Antisense deoxyoligonucleotides that trigger RNase H activity can thus be used as cancer therapeutic agents (Crooke, 1996; Curcio et al., 1997).

[0225] RNase H-related sequences can possess or interact with rnaseH, Gag_p30, rvt, and rve domains, all of which are described above.

SH3-Related Sequences

[0226] Src homology region 3 (SH3) is a polypeptide domain commonly found in intracellular signaling proteins; it binds with moderate affinity and selectivity to proline-rich ligands. SH3 domains are heterogeneous; different SH3 domains bind to different proline-rich sequences (Gmeiner and Horita, 2001). SH3 domains are involved in a wide variety of biological processes, including mediating the assembly of large multiprotein complexes, regulating enzyme activity, and modulating the local concentration or subcellular localization of signaling pathway components (Mayer, 2001). Examples of SH3-related sequences include phosphotyrosine receptors, membrane associated guanylate kinases, mitogen-activated protein kinases, myosin 1,

the Crk adaptor protein, phospholipase C-γ, Grb2, Sos, src-SH3, Abl-SH3, the Nck adaptor, and alpha-spectrin-SH3.

[0227] SH3-related sequences can possess or interact with SH3 domains, which are protein domains of approximately 50-70 amino acids, and are present in a large number of proteins involved in intracellular signaling (http://pfam.wustl.edu/cgi-bin/getdesc?name=SH3). SH3-related sequences can also possess or interact with SH3 domain-binding protein 5 (SH3BP5) domains, which are protein domains that act as a substrate for c-Jun N-terminal kinase (http://pfam.wustl.edu/cgi-bin/getdesc?name=SH3BP5).

Stem Cell-Related Sequences

[0228] Stem cells are pluripotent or multipotent cells that generate maturing cells in multiple differentiation lineages. Pluripotent cells have the capacity to differentiate into each and every cell present in the organism. Embryonic stem cells are pluripotent; they can differentiate into any of the cells present in the adult. Multipotent cells have the ability to differentiate into more than one cell type. Organ-specific stem cells are multipotent; they can differentiate into any of the cells of the organ they inhabit.

[0229] When they divide *in vivo*, both pluripotent and multipotent stem cells can maintain their pluripotency or multipotency while giving rise to differentiated progeny. Thus, stem cells can produce replicas of themselves which are pluri- or multipotent, and are also able to differentiate into lineage-restricted committed progenitor cells. For example, hematopoeitic stem cells, which are multipotent cells specifically able to form blood cells, can divide to produce replicate hematopoeitic stem cells. They can also divide to produce more highly differentiated cells, which are precursors of blood cells. The precursors differentiate, sometimes through several generations of cells, into blood cells. A hematopoetic stem cell can also divide into a cell with the capacity to form, for example, a relatively undifferentiated cell that is committed to differentiate into, i.e., granulocytes, or erythrocytes, or another type of blood cell.

[0230] Stem cells can also reproduce and differentiate *in vitro*. Embryonic stem cells have been directed to differentiate into cardiac muscle cells *in vitro* and, alternatively, into early progenitors of neural stem cells, and then into mature neurons and glial cells in vitro (Trounson, 2002).

[0231] Stem cell therapy is effective in treating cancer in humans (Slavin et al., 2001), and offers several advantages over traditional cancer therapies (Weissman, 2000). One advantage of stem cell therapy exists when used in conjunction with radiation therapy. In radiation therapy for cancer, the dose of radiation necessary to kill the cancer cells in an organ can also be sufficient to destroy the healthy cells of the organ. In combined stem cell and radiation therapy, an organ is first treated with sufficient radiation to destroy all of the cancer cells and most or all of the healthy cells, but then stem cells are infused to repopulate the organ. In the ensuing weeks, as the cancer cells and healthy cells die, the stem cells replace the healthy cells. Another advantage of this approach, compared to heterologous organ transplants, is that there is no risk of rejection, since stem cells do not provoke an immune response. A further advantage is that stem cells are inherently programmed to regulate their numbers and differentiation status, i.e., once provided to the patient, the necessary number will differentiate, and the rest will remain undifferentiated (Weissman, 2000).

[0232] Stem cell therapy is also effective in treating autoimmune disease in humans. For example, immunosuppression in conjunction with stem-cell transplantation has induced remission in patients with refractory, severe rheumatic autoimmune disease (Van Laar and Tyndall, 2003). Patients with rheumatoid arthritis, systemic lupus erythematosus, systemic sclerosis, and juvenile idiopathic arthritis have benefited from stem cell transplants (Van Laar and Tyndall, 2003).

[0233] Preclinical studies also suggest the potential of stem cell transplantation for the treatment of neural and muscular injuries and disorders, including those of the central nervous system, peripheral nervous system, and skeletal, cardiac and smooth muscle (Deasy and Huard, 2002). Stem cells transplanted into the bone marrow of mice migrate to the site of injured muscle and differentiate into new muscle cells. For example, patients with myasthenia gravis, muscular dystrophies, amyotrophic lateral sclerosis, congestive heart failure, Parkinson's disease, and Alzheimer's disease may benefit from stem cell therapy (Henningson, 2003).

[0234] In addition to therapeutic uses, research using stem cells can provide useful information about normal stem cell function and the pathogenesis of disease. Stem cells derived from a patient with a genetic disease can provide a tool for studying that disease. To derive these stem cells, a somatic cell, i.e., a cell that is not in the oocyte or spermatocyte lineage, is donated by the patient, and the nucleus is removed and transferred to an unfertilized human oocyte. This nuclear transplant

procedure produces, at the blastocyst stage of development, embryonic stem cells with the same set of genes as the patient with the genetic disease. Studying these cells, and their progeny in vitro, permits analysis of a specific model of the disease. For example, placing stem cells derived from a patient with a genetic disorder under the control of various stem cell regulatory factors can elicit abnormal responses from the affected stem cells compared to stem cells derived from a healthy individual's somatic nucleus.

[0235] Embryonic stem cell-related sequences can possess or interact with the stem cell factor (SCF) domain, a transmembrane domain having a soluble, secreted form, which is involved in hematopoeisis, and which binds to and activates a receptor tyrosine kinase, stimulating the proliferation of mast cells and augmenting the proliferation of myeloid and lymphoid hematopoietic progenitors in bone marrow culture (http://pfam.wustl.edu/cgi-bin/getdesc?name=SCF).

[0236] Certain stem cell related sequences can possess the ability to maintain the stem cell in undifferentiated state while allowing cell proliferation. Such compositions can be useful in *ex vivo* cell therapy to expand populations of cells for cell replacement therapy.

[0237] Certain stem cell related sequences can possess the ability to cause cell differentiation to a relatively mature cell type and are useful to *in vivo* or *ex vivo* therapy to compensate for deficiency of such relatively mature cell type.

Synthetase-Related Sequences

[0238] A synthetase is an enzyme that catalyzes the synthesis of a molecule. Synthetases comprise a broad class of enzymes; they catalyze the synthesis of nucleic acids, peptides, and lipids (Agou et al., 1996). Examples of synthetases include lysyl-tRNA synthetase, asparaginyl t-RNA synthetase, holocarboxylase synthetase, carbamyl phosphate synthetase I, and argininosuccinate synthetase.

[0239] Synthetase-related sequences can possess or interact with transfer RNA synthetase domains, which are protein domains that activate amino acids and transfer them to specific transfer RNA molecules as a step in protein biosynthesis (http://pfam.wustl.edu/cgi-bin/getdesc?name=tRNA-synt_2). The 20 aminoacyl-tRNA synthetases are divided into class I and class II, each of which contain multiple synthetases with different specificities. For example, there is a protein domain involved in the asparagines, aspartic acid, and lysine synthesis (http://pfam.wustl.edu/cgi-bin/textsearch?terms=trna-synt&search_what=all& sections=

DE&sections=CC&size=100). Synthetase-related sequences can also possess or interact with lipid-A-disaccharide synthetase (LpxB) domains, which are protein domains that catalyze the synthesis of disaccharides (http://pfam.wustl.edu/cgi-bin/getdesc? name=LpxB).

TATA Box-Related Sequences

[0240] A TATA box is a consensus sequence in the promoter region of many eucaryotic genes that binds a general transcription factor and plays a role in specifying the position for transcription initiation. TATA boxes are generally found approximately 25 nucleotides before the site of transcription initiation (Chalut et al., 1995). Examples of TATA box-related sequences include TATA box binding protein, 13 TATA/TBP, and small nuclear RNA-activating protein 190 Myb DNA.

[0241] TATA box-related sequences can possess or interact with transcription factor TFIID, also known as the TATA-binding protein (TBP) domain, which is a protein domain that specifically binds to the TATA box promoter element (http://pfam.wustl.edu/cgi-bin/getdesc?name=TBP). TATA box-related sequences can also possess or interact with HMG14 and HMG17 (HMG14_17) domains, which are members of a family of high mobility group proteins, described above (http://pfam.wustl. edu/cgi-bin/getdesc? name=HMG14_17).

Tat-Related Sequences

[0242] Tat is a human immunodeficiency virus (HIV) protein involved in viral production of new RNA genomes and new complete viral particles. Tat is also involved in AIDS pathogenesis; it plays a role in reactivating latent viruses, e.g., the JC retrovirus; it is involved in the development of AIDS-related Kaposi's Sarcoma; and it depresses the function of, and induces apoptosis in, helper CD4 cells (Yu et al., 1995). Examples of Tat-related sequences include Tat-associated proteins, e.g., Tap, HIV-1 Rev, and tat-associated kinase (also known as positive transcriptional elongation factor b).

[0243] Tat-related sequences can possess or interact with transactivating regulatory protein (Tat) domains, which are protein domains that contribute to efficient transcription of a viral genome (http://pfam.wustl.edu/cgi-bin/getdesc?name=Tat). Tat-related sequences can also possess or interact with mitochondrial glycoprotein (MAM33) domains, which are protein domains found in mitochondrial matrix proteins, and which can be involved in mitochondrial oxidative

phosphorylation and in interactions between the nucleus and the mitochondria (http://pfam.wustl.edu/cgi-bin/getdesc?name=MAM33).

Transferase-Related Sequences

[0244] Transferases are enzymes that transfer a designated group of atoms from a donor molecule to an acceptor molecule. For example, acyl transferases transfer acyl groups, methyl transferases transfer methyl groups, nucleotidyl transferases transfer nucleotides, prenyltransferases transfer prenyl groups, and glycosyl transferases transfer glycosyl groups (Lin et al., 1996). Examples of transferases include acetyltransferases, hydroxymethyltransferases, sialyltransferases, arginine N-methyltransferase, glucoronosyltransferase, NTP-transferase, and GDP-mannose pyrophosphorylase B.

[0245] Transferase-related sequences possess or interact with UDP-glucuronosyl and UDP-glucosyl transferase domains, which are protein domains found in a superfamily of enzymes that catalyze the addition of the glycosyl group from a UTP-sugar to a small hydrophobic molecule (http://pfam.wustl.edu/cgi-bin/getdesc?name=UDPGT). Transferase-related sequences also possess or interact with nucleotide transferase (NTP_transferase) domains, which are protein domains that transfer nucleotides onto phosphorylated sugars (http://pfam.wustl.edu/cgi-bin/getdesc?name=NTP_transferase).

Transposase-Related Sequences

[0246] Transposases are site-specific recombination enzymes that catalyze the transposition of a segment of DNA from one part of the genome to another. The movable segments are called transposable elements; each transposable element is occasionally moved by a transposase, which functions as an integrase, by inserting DNA sequences into other DNA sequences. Transposases are often encoded by the DNA of the transposable element itself. Transposases bind specifically to terminal inverted repeats of 10-500 bp that are characteristically part of transposable elements (Smit and Riggs, 1996). They catalyze both cutting and pasting of a transposable element from one segment of the genome to another. Sequences related to transposases can have other functions, e.g., as transcription factors, or in the assembly of centromere proteins (Smit and Riggs, 1996). Examples of transposase-related sequences include mariner, pogo, hobo, tigger, MER37, Galileo, Occan, Impala, Tn MERI1, MsqTc3, and the sleeping beauty transposon system (Robertson and Zumpano, 1997; Robertson, 1996; Smit and Riggs, 1996).

[0247] Transposase-related sequences can possess or interact with a transposase 1 (Transposase_1) domain, which is characterized by sequences that can excise and/or insert mobile genetic elements such as transposons or insertion sequences; for example, mariner possesses a transposase 1 domain (http://pfam.wustl.edu/cgi-bin/getdesc? name= Transposase_1). Transposase-related sequences can also possess or interact with L1 transposable element (Transposase_22) domains, which have been described above. Transposase-related sequences can also possess or interact with a DDE endonuclease (DDE) domain, which is responsible for coordinating metal ions needed for endonuclease catalytic activity (http://pfam.wustl.edu/cgi-bin/getdesc? name=DDE). Transposase-related sequences can additionally possess or interact with a zinc finger, C2H2 type (zf-C2H2) domain, which bind nucleic acids using a mechanism that involves coordinating a zinc atom with a pair of cysteine residues and a pair of histidine residues (http://pfam.wustl.edu/cgi-bin/getdesc?name=zf-C2H2). Transposase-related sequences can also possess or interact with a reverse transcriptase (rvt) domain, and/or a low-density lipoprotein receptor (ldl_rece) domain, both of which are described above.

Ubiquitin-Related Sequences

[0248] Ubiquitin is a protein found in all eucaryotic cells examined to date. When it is linked to the lysine side chain of a protein by the formation of an amide bond with its C-terminal glycine, ubiquitin renders the ubiquitin-bound protein subject to rapid proteolysis in the proteasome. In addition to its role in the selective degradation of cellular proteins, ubiquitin also plays a role in maintaining chromosome structure, regulating gene expression, responding to stresses on the organism, the regulation of gene expression, and ribosome biogenesis. Examples of ubiquitin-related sequences include elongins, ubiquitin-specific proteases, ubiquitin-calmodulin ligase, ubiquitin carrier protein kinase, ubiquitin N-alpha-protein hydrolase, and the small ubiquitin-related modifier (Sumo-1) (Kamitani et al., 1997).

[0249] Ubiquitin-related sequences can possess or interact with a ubiquitin domain, which is a conserved sequence of approximately 76 amino acid residues that comprise the protein ubiquitin (http://pfam.wustl.edu/cgi-bin/getdesc?name=ubiquitin). Ubiquitin-related sequences can also possess or interact a ubiquitin carboxyl-terminal hydrolase (UCH) domain, which is a protein domain that comprises a thiol protease that recognizes and hydrolyses the peptide

bond at the C-terminal glycine of ubiquitin (http://pfam.wustl.edu/cgi-bin/get desc?name=UCH).

Virus-Related Sequences

[0250] The human chromosome has integrated endogenous genes that are related to viral genes. Some endogenous viral genes, e.g., the retroviral HERV-W family, are widely and heterogeneously dispersed among human chromosomes (Voisset et al., 2000; Everett et al., 1997; Werner et al., 1990). Endogenous proviruses are usually transcriptionally silent, but are expressed under certain conditions (Coffin et al., 1997). Endogenous viral expression can be specific to host factors, such as cell type or stage of differentiation, as well as other factors including the position on the chromosome, the influence of *cis*-acting sequences, or the presence of host-mediated DNA methylation (Coffin).

[0251] Endogenous viral expression can have a number of consequences, both beneficial and detrimental. Among the beneficial consequences is the ability of endogenous retroviruses to confer resistance to infection by exogenous viruses. For example, mice with endogenous mouse mammary tumor virus (MMTV) can be immune to exogenous infection (Golovkina, et al., 1992). Among the detrimental effects is a causative role in disease. Evidence indicates an association between endogenous viruses with cancers and autoimmune diseases (Coffin et al., 1997). For example, spontaneous tumors of specific origin, murine mammary adenocarcinomas, and murine T-cell lymphomas have been associated with the presence of specific endogenous retroviruses. Furthermore, a transformed phenotype is associated with the increased transcription of certain classes of endogenous viral elements (Coffin et al., 1997). With respect to autoimmune disease, an endogenous virus that influences the immunoregulatory process has been associated with spontaneous autoimmune thyroiditis in a chicken model of human Hashimoto disease (Wick et al., 1987). Examples of viral-related proteins include hepatitis B virus x-interacting protein, herpesvirus associated ubiquitin-specific protease, and Coxsackievirus and adenovirus receptor precursor.

[0252] Viral-related sequences can possess or interact with rvt, rve, and gag_p30 sequences, all of which are described above.

Zinc Finger-Related Sequences

[0253] A zinc finger domain is a small, self-folding, structural motif of 25 to 30 amino-acid residues present in many nucleic acid-binding proteins. It is comprised

of a polypeptide loop held in a hairpin bend and bound to a zinc atom, and includes two conserved cysteine and two conserved histidine residues. Many classes of zinc fingers have been characterized according to the number and positions of the conserved histidine and cysteine residues. The amino acid configuration that holds the zinc atom in a tetrahedral array has a finger-like projection that interacts with nucleotides in the major groove of the bound nucleic acid. Zinc finger motifs have conserved regions near the zinc molecule, and variable regions at the nucleic acid binding site that provide specificity for the nucleic acid sequences they bind. Zinc finger proteins have a variety of functions, including as transcription regulators and intracellular receptors. Zinc finger domains are also involved in protein-protein interactions, e.g., those involving protein kinase C. Recently, zinc finger nucleases have been used to target genes for gene replacement by homologous recombination (Bibikova et al., 2003). Examples of zinc finger proteins include XC3H-3b, the transcription factor Slug, and transcription factor IIIA.

[0254] Zinc finger-related sequences can possess or interact with a zinc finger C2H2 type (zf-C2H2) domain, which binds a zinc atom with two cysteine and two histidine residues, and is utilized, e.g., in RNA transcription (http://pfam.wustl.edu/cgi-bin/getdesc?name=zf-C2H2). Zinc finger-related sequences can also possess or interact with a C3HC4 type, RING finger (zf-C3HC4) domain, which is a specialized type of zinc finger domain comprised of 40 to 60 amino acids that binds two zinc atoms; variants of RING-finger domains include the C3HC4-type and the C3H2C3-type (http://pfam.wustl.edu/cgi-bin/getdesc?name=zf-C3HC4). Proteins with RING-finger domains have developmental and functional roles; they are involved in intracellular receptor binding, and in mediating protein-protein interactions (Gray et al., 2000). RING-finger domains can exhibit ubiquitin-protein ligase activity, and can bind to E2 ubiquitin-conjugating enzymes.

[0255] Zinc finger-related sequences can also possess or interact with a zinc knuckle (zf-CCHC) domain, which is an 18-amino acid zinc finger domain found in RNA-binding and single strand DNA-binding proteins; they are often involved in eukaryotic gene regulation (http://pfam.wustl.edu/cgi-bin/getdesc?name=zf-CCHC). Zinc knuckles are also found in retroviral gag and nucleocapsid proteins, where they function in genome packaging, and early in the infection process. Zinc finger-related sequences can also possess or interact with a BTB/POZ (BTB) domain, which mediates both homomeric and heteromeric protein dimerization (http://pfam.wustl.

edu/cgi-bin/getdesc?name=BTB). Zinc finger-related sequences can also possess or interact with NF-X1 type zinc finger (zf-NF-X1) domains, which are found in the transcriptional repressor NK-X1, where they repress transcription of HLA-DRA, and in the shuttle craft protein, which plays a role in late stage embryonic neurogenesis (http://pfam.wustl.edu/cgi-bin/getdesc?name=zf-NF-X1). Zinc finger-related sequences can also possess or interact with a KRAB box (KRAB) domain, also known as a Kruppel-associated box, which is comprised of approximately 75 amino acids, enriched in charged amino acids, and involved in protein-protein interactions (http://pfam.wustl.edu/cgi-bin/getdesc? name=KRAB). KRAB domains can function as transcription factors, e.g., as a transcriptional repressor, and can assume roles in cell differentiation and development (Aubry et al., 1992; Lovering and Trowsdale, 1991). Zinc finger-related sequences can possess or interact with a transposase_22 domain, which is described above.

INDUSTRIAL APPLICABILITY

[0256] The invention provides sequences related to secreted sequences, single-transmembrane sequences, multiple-transmembrane sequences, kinase-related sequences, ligase-related sequences, nuclear hormone receptor-related sequences, phosphatase-related sequences, protease-related sequences, phosphodiesterase-related sequences, kinesin-related sequences, immunoglobulin-related sequences, T-cell receptor-related sequences, glycosylphosphatidylinositol anchor-related sequences, and sequences related to other nucleic acid and amino acid sequences of the invention, including activators, adaptors, adhesion molecules, ATPases, ATP, breakpoints, channels, checkpoints, complexes, dehydrogenases, disintegrins, endopeptidases, germ-cells, GTPases, helicases, hydrolases, integrases, integrins, isomerases, membranes, mucins, oxygenases, peroxidases, phopholipases, prosaposins, proteosomes, reductases, reverse trancriptases, RNases, RNases H, SH3, synthetases, TATA boxes, Tat proteins, transferases, transposases, ubiquitins, and viruses. The invention provides for novel polynucleotides, related novel polypeptides and active fragments thereof, as well as novel nucleic acid compositions encoding these polypeptides, compositions comprising the related polypeptides, and methods for their use.

[0257] The present invention also provides for vectors, host cells, and methods for producing the polynucleotides and polypeptides of the invention in these

vectors and host cells. The present invention further provides for antisense molecules that are capable of regulating the expression of the polynucleotides or polypeptides herein. In addition, modulators, including antibodies that bind specifically to the polypeptides or modulate the activity of the polypeptides, are also provided.

[0258] The present polynucleotides, polypeptides, and modulators find use in therapeutic agent screening/discovery applications, such as screening for receptors or competitive ligands, for use, for example, as small molecule therapeutic drugs. Also provided are methods of modulating a biological activity of a polypeptide and methods of treating associated disease conditions, particularly by administering modulators of the present polypeptides, such as small molecule modulators, antisense molecules, and specific antibodies.

[0259] The present polypeptides, polynucleotides, and modulators find use in a number of diagnostic, prophylactic, and therapeutic applications. The polynucleotides and polypeptides of the invention can be detected by methods provided herein; these methods are useful in diagnosis, and can be accomplished by the use of diagnostic kits. The polynucleotides and polypeptides of the invention are useful for treating a variety of disorders, including cancer, proliferative disorders, inflammatory disorders, immune disorders, viral disorders, and other metabolic disorders. For example, subjects who suffer from a deficiency, or a lack of a particular protein, or are otherwise in need of such protein to repair or enhance a desirable function, benefit from the administration of a protein or an active fragment thereof by any conventional routes of administration. These include therapeutic vaccines in the form of nucleic acid or polypeptide vaccines, such as cancer vaccines, where the vaccines can be administered alone, such as naked DNA, or can be facilitated, such as via viral vectors, microsomes, or liposomes. Therapeutics antibodies include those that are administered alone or in combination with cytotoxic agents, such as radioactive or chemotherapeutic agents.

[0260] In particular, the polypeptides, polynucleotides, and modulators of the present invention can be used to treat cancers, including, but not limited to, cancers of the prostate, breast, bone, soft tissue, liver, kidney, ovary, cervix, skin, pancreas, and brain, as well as leukemias, lymphomas, lung cancers such as adenocarcinomas and squamous cell carcinoma, and cancers of gastrointestinal organs such as stomach, colon, and rectum. Further, the polypeptides, polynucleotides, and modulators of the present invention can be used to treat inflammatory, immune,

bacterial, viral, and metabolic diseases, disorders, syndromes, or conditions, including, but not limited to, intestinal inflammation and immunity, autoimmune thyroiditis, and retroviral infections, as well as tissue and/or organ hypertrophy.

DISCLOSURE OF THE INVENTION

[0261] The present invention features an isolated polynucleotide that encodes a polypeptide. In some embodiments, the polypeptide has at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 97%, at least about 98%, or at least about 99% amino acid sequence identity with an amino acid sequence derived from a polynucleotide sequence chosen from at least one nucleotide sequence according to SEQ ID NOS.: 1 - 209 and 419 - 627. In some embodiments, the polypeptide has an amino acid sequence chosen from at least one amino acid sequence according to SEQ ID. NOS. 210 - 418. In many embodiments, the polypeptide has at least one activity associated with the naturally occurring encoded polypeptide.

[0262] In some embodiments, the polypeptide includes a signal peptide. In alternative embodiments, the polypeptide comprises a mature form of a protein, from which the signal peptide has been cleaved. In other embodiments, the polypeptide is a signal peptide. In a further aspect, the invention provides fragments of a polypeptide chosen from at least one amino acid sequence according to SEQ ID NOS.: 210 - 418, where each fragment is an extracellular fragment of the polypeptide, or an extracellular fragment of the polypeptide minus the signal peptide. The invention provides an N-terminal fragment containing a Pfam domain and a C-terminal fragment containing a Pfam domain and either or both may be biologically active.

[0263] In yet other embodiments, the polypeptides function as secreted proteins. In yet further embodiments, the polypeptides function as single-transmembrane proteins. In yet further embodiments, the polypeptides function as multiple-transmembrane proteins. In yet further embodiments, the polypeptides function as kinases. In yet further embodiments, the polypeptides function as protein kinases. In yet further embodiments, the polypeptides function as ligases. In yet further embodiments, the polypeptides function as nuclear hormone receptors. In yet further embodiments, the polypeptides function as phosphatases. In yet further embodiments, the polypeptides function as proteases. In yet further embodiments, the polypeptides function as phosphodiesterases. In yet further embodiments, the

polypeptides function as kinesins. In yet further embodiments, the polypeptides function as immunoglobulins. In yet further embodiments, the polypeptides function as T-cell receptors. In yet further embodiments, the polypeptides function as glycosylphosphatidylinositol anchors.

[0264] In yet further embodiments, the polypeptides function as cytokines. In still further embodiments, the polypeptides function as immune cells. In further embodiments, the polypeptides function as antigens. In yet further embodiments, the polypeptides function as receptors. In other embodiments, the polypeptides function as binding proteins. In other embodiments, the polypeptides function as factors. In further embodiments, the polypeptides function as growth factors. In further embodiments, the polypeptides function as heat-shock proteins. In some embodiments, the polypeptides function as membrane transport proteins. In yet further embodiments, the polypeptides function as ribosomal proteins. In some embodiments, the polypeptides function as zinc fingers. In some embodiments, the polypeptides function as embryonic stem cell-related peptides. In still further embodiments, the polypeptides function in pathological states. In other embodiments, the polypeptides function as one or more of these.

[0265] In yet further embodiments, the polypeptides function as activators. In yet further embodiments, the polypeptides function as adaptors. In yet further embodiments, the polypeptides function as adhesion molecules. In yet further embodiments, the polypeptides function as ATPases. In yet further embodiments, the polypeptides function as ATP-related polypeptides. In further embodiments, the polypeptides function as channel-related polypeptides. In yet further embodiments, the polypeptides function as checkpoint-related polypeptides. In yet further embodiments, the polypeptides function as complexes. In yet further embodiments, the polypeptides function as dehydrogenases. In yet further embodiments, the polypeptides function as disintegrins. In yet further embodiments, the polypeptides function as endopeptidases. In yet further embodiments, the polypeptides function as germ-cells. In yet further embodiments, the polypeptides function as GTPases. In yet further embodiments, the polypeptides function as helicases. In yet further embodiments, the polypeptides function as hydrolases. In yet further embodiments, the polypeptides function as integrases. In yet further embodiments, the polypeptides function as integrins. In yet further embodiments, the polypeptides function as isomerases. In yet further embodiments, the polypeptides function as membranes. In

yet further embodiments, the polypeptides function as mucins. In yet further embodiments, the polypeptides function as oxygenases. In yet further embodiments, the polypeptides function as peroxidases. In some embodiments, the polypeptides function as phospholipases. In yet further embodiments, the polypeptides function as prosaposins. In yet further embodiments, the polypeptides function as proteasomes. In yet further embodiments, the polypeptides function as reductases. In other embodiments, the polypeptides function as reverse transcriptase-related polypeptides. In yet further embodiments, the polypeptides function as RNases. In further embodiments, the polypeptides function as RNase H-related polypeptides. In yet further embodiments, the polypeptides function as SH3-related polypeptides. In yet further embodiments, the polypeptides function as synthetases. In yet further embodiments, the polypeptides function as TATA box-related polypeptides. In yet further embodiments, the polypeptides function as TAT-related polypeptides. In yet further embodiments, the polypeptides function as transferases. In yet further embodiments, the polypeptides function as transposases. In yet further embodiments, the polypeptides function as ubiquitin-related polypeptides. In yet further embodiments, the polypeptides function as virus-related polypeptides. In other embodiments, the polypeptides function as one or more of these.

[0266] The present invention features an isolated polynucleotide that hybridizes under stringent hybridization conditions to a coding region of at least one nucleotide sequence shown in SEQ ID NOS.: 1 - 209, 419 - 627, or a complement thereof.

[0267] The present invention features an isolated polynucleotide that shares at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 97%, at least about 98%, at least about 99% nucleotide sequence identity with a nucleotide sequence of the coding region of at least one sequence shown in SEQ ID NOS.: 1 - 209, 419 - 627, or a complement thereof. In some embodiments, a subject polynucleotide has the nucleotide sequence shown in at least one of SEQ ID NOS.: 1 - 209, 419 - 627, or a coding region thereof.

[0268] The present invention also features a vector, e.g., a recombinant vector, that includes a subject polynucleotide, and a promoter the drives its expression. This vector can transform a host cell, and the present invention further features such host cells, e.g., isolated *in vitro* host cells, and *in vivo* host cells, that comprise a polynucleotide of the invention, or a recombinant vector of the invention.

[0269] The present invention further features a library of polynucleotides, wherein at least one of the polynucleotides comprises the sequence information of a polynucleotide of the invention. In specific embodiments, the library is provided on a nucleic acid array. In some embodiments, the library is provided in computer-readable format.

[0270] The present invention features a pair of isolated nucleic acid molecules, each from about 10 to about 200 nucleotides in length. The first nucleic acid molecule of the pair comprises a sequence of at least 10 contiguous nucleotides having 100% sequence identity to at least one nucleic acid sequence shown in SEQ ID NOS.: 1 - 209 and 419 - 627. The second nucleic acid molecule of the pair comprises a sequence of at least 10 contiguous nucleotides having 100% sequence identity to the reverse complement of at least one nucleic acid sequence shown in SEQ ID NOS.: 1 - 209 and 419 - 627. The sequence of said second nucleic acid molecule is located 3' of the nucleic acid sequence of the first nucleic acid molecule shown in SEQ ID NOS.: 1 - 209 and 419 - 627. The pair of isolated nucleic acid molecules are useful in a polymerase chain reaction or in any other method known in the art to amplify a nucleic acid that has sequence identity to the sequences shown in SEQ ID NOS.: 1 - 209 and 419 - 627, particularly when cDNA is used as a template.

[0271] The invention features a method of determining the presence of a polynucleotide substantially identical to a polynucleotide sequence shown in the Sequence Listing, or a complement of such a nucleotide by providing its complement, allowing the polynucleotides to interact, and determining whether such interaction has occurred.

[0272] The invention further features methods of regulating the expression of the subject polynucleotides and encoded polypeptides. The invention provides a method of inhibiting transcription or translation of a first polynucleotide encoding a first polypeptide of the invention by providing a second polynucleotide that hybridizes to the first polynucleotide, and allowing the first polynucleotide to contact and bind to the second polynucleotide. The second polynucleotide can be chosen from an antisense molecule, a ribozyme, and an interfering RNA (RNAi) molecule.

[0273] The present invention further features an isolated polypeptide, e.g., an isolated polypeptide encoded by a polynucleotide, and biologically active fragments of such polypeptide. In some embodiments, the polypeptide is a fusion protein. In some embodiments, the polypeptide has one or more amino acid substitutions, and/or

insertions and/or deletions, compared with at least one sequence shown in SEQ ID NOS.: 210 - 418. In some embodiments, the polypeptide has an amino acid sequence derived from at least one nucleotide sequence shown in SEQ ID NOS.: 1 - 209 and 419 - 627. In some embodiments, the polypeptide has an amino acid sequence substantially identical to at least one sequence shown in SEQ ID NOS.: 210 - 418.

[0274] The invention also provides a method of making a polypeptide of the invention by providing a nucleic acid molecule that comprises a polynucleotide sequence encoding a polypeptide of the invention, introducing the nucleic acid molecule into an expression system, and allowing the polypeptide to be produced.

[0275] In some embodiments, the method involves *in vitro* cell-free transcription and/or translation. For example, the expression system can comprise a cell-free expression system, such as an *E. coli* system, a wheat germ extract system, a rabbit reticulocyte system, or a frog oocyte system.

[0276] In certain other embodiments, the expression system can comprise a prokaryotic or eukaryotic cell, for example, a bacterial cell expression system, a fungal cell expression system, such as yeast or *Aspergillus*, a plant cell expression system, e.g., a cereal plant, a tobacco plant, a tomato plant, or other edible plant, an insect cell expression system, such as SF9 of High Five cells, an amphibian cell expression system, a reptile cell expression system, a crustacean cell expression system, an avian cell expression system, a fish cell expression system, or a mammalian cell expression system, such as one using Chinese Hamster Ovary (CHO) cells. In some embodiments, the method involves culturing a subject host cell under conditions such that the subject polypeptide is produced by the host cells; and recovering the subject polypeptide from the culture, e.g., from within the host cells, or from the culture medium. In further embodiments, the polypeptide can be produced *in vivo* in a multicellular animal or plant, comprising a polynucleotide encoding the subject polypeptide.

[0277] The present invention further features a non-human animal injected with at least one polynucleotide comprising at least one nucleotide sequence chosen from SEQ ID NOS.: 1 - 209 and 419 - 627, and/or at least one polypeptide comprising at least one amino acid sequence chosen form SEQ ID NOS.: 210 - 418.

[0278] The invention further provides a kit comprising one or more of a polynucleotide or polypeptide, which may include instructions for its use. Such kits

are useful in diagnostic applications, for example, to detect the presence and/or level of a polypeptide in a biological sample.

MODES FOR CARRYING OUT THE INVENTION

Brief Description of the Tables

[0279] Each sequence shown in Tables 1-3 is identified by a Five Prime Therapeutics, Inc. (FP) identification number (FP ID). Table 1 specifies the predicted number of amino acid residues in each FP protein of the invention (Length, Predicted Protein). Table 1 also specifies the percent of the FP sequence that is covered by the public National Center for Information Biotechnology (NCBI) database (Prediction Covered by Public). Table 1 also describes the characteristics of the protein in the NCBI database displaying the greatest degree of similarity to each claimed sequence. This protein is described by its NCBI accession number (Top Hit Accession No.), and by the NCBI's annotation of that sequence (Top Hit Annotation).

[0280] Table 2 describes the characteristics of the human protein in the NCBI database with the greatest degree of similarity to each claimed sequence. The predicted number of amino acids of this human protein is specified (Length, Human Top Hit). Table 2 also specifies any existing protein family (Pfam) classification for these human sequences. Table 2 specifies the result of the algorithm described above that predicts whether the claimed FP sequence is secreted (Tree Vote, Secreted). Table 2 sets forth the position of the amino acid residues comprising the signal peptide sequences (SP Positions) of the claimed FP sequences. Table 2 also specifies the position(s), if any, of the amino acid residues comprising the transmembrane domains in each claimed FP sequence (TM domains), and the number of transmembrane domains of each claimed FP sequence (TM Total).

[0281] Table 3 describes the characteristics of the Fantom mouse protein with the greatest degree of similarity to the claimed sequences. The Fantom database was compiled by the Fantom Consortium and is accessible, for example, at http://fantom.gsc.riken.go.jp/db/ (Bono et al., 2002). It provides curated functional annotation to full-length mouse sequences (Okzaki et al., 2002). The similarities of the claimed sequences of the invention with the annotated sequences in Tables 1-3 suggest that they may share structural and functional properties, and exhibit similar expression profiles and localizations.

Definitions

[0282] "Related sequences" include nucleotide and amino acid sequences that are involved in the function of their referent. For example, "receptor-related sequences" include all sequences that are involved in receptor function. This includes, but is not limited to, sequences that are involved in receptor synthesis, receptor regulation, receptor effector function, and receptor degradation. "Related sequences" also encompass complementary nucleic acid sequences, and biologically active fragments of nucleic acid and amino acid sequences.

[0283] The terms "polynucleotide," "nucleotide," "nucleic acid," "polynucleic molecule," "nucleotide molecule," "nucleic acid molecule," "nucleic acid sequence," "polynucleotide sequence," and "nucleotide sequence" are used interchangeably herein to refer to polymeric forms of nucleotides of any length. The polynucleotides can contain deoxyribonucleotides, ribonucleotides, and/or their analogs or derivatives. For example, nucleic acids can be naturally occurring DNA or RNA, or can be synthetic analogs, as known in the art. The terms also encompass genomic DNA, genes, gene fragments, exons, introns, regulatory sequences or regulatory elements (such as promoters, enhancers, initiation and termination regions, other control regions, expression regulatory factors, and expression controls), DNA comprising one or more single-nucleotide polymorphisms (SNPs), allelic variants, isolated DNA of any sequence, and cDNA. The terms also encompass mRNA, tRNA, rRNA, ribozymes, splice variants, antisense RNA, antisense conjugates, RNAi, and isolated RNA of any sequence. The terms also encompass recombinant polynucleotides, heterologous polynucleotides, branched polynucleotides, labeled polynucleotides, hybrid DNA/RNA, polynucleotide constructs, vectors comprising the subject nucleic acids, nucleic acid probes, primers, and primer pairs. The polynucleotides can comprise modified nucleic acid molecules, with alterations in the backbone, sugars, or heterocyclic bases, such as methylated nucleic acid molecules, peptide nucleic acids, and nucleic acid molecule analogs, which may be suitable as, for example, probes if they demonstrate superior stability and/or binding affinity under assay conditions. Analogs of purines and pyrimidines, including radiolabeled and fluorescent analogs, are known in the art. The polynucleotides can have any three-dimensional structure, and can perform any function, known or as yet unknown. The terms also encompass single-stranded, double-stranded and triple helical molecules that are either DNA, RNA, or hybrid DNA/RNA and that may encode a

full-length gene or a biologically active fragment thereof. Biologically active fragments of polynucleotides can encode the polypeptides herein, as well as anti-sense and RNAi molecules. Thus, the full length polynucleotides herein may be treated with enzymes, such as Dicer, to generate a library of short RNAi fragments which are within the scope of the present invention.

[0284] The novel polynucleotides herein include those shown in the Tables, SEQ ID NOS.: 1 - 209 and 419 - 627, as well as those that encode the polypeptides of SEQ ID NOS.: 210 - 418, and biologically active fragments thereof. The polynucleotides also include modified, labeled, and degenerate variants of the nucleic acid sequences, as well as nucleic acid sequences that are substantially similar or homologous to nucleic acids encoding the subject proteins.

[0285] A "biologically active" entity, or an entity having "biological activity," is one having structural, regulatory, or biochemical functions of a naturally occurring molecule or any function related to or associated with a metabolic or physiological process. Biologically active polynucleotide fragments are those exhibiting activity similar, but not necessarily identical, to an activity of a polynucleotide of the present invention. The biological activity can include an improved desired activity, or a decreased undesirable activity. For example, an entity demonstrates biological activity when it participates in a molecular interaction with another molecule, or when it has therapeutic value in alleviating a disease condition, or when it has prophylactic value in inducing an immune response to the molecule, or when it has diagnostic value in determining the presence of the molecule, such as a biologically active fragment of a polynucleotide that can be detected as unique for the polynucleotide molecule, or that can be used as a primer in PCR.

[0286] The term "degenerate variant" of a nucleic acid sequence refers to all nucleic acid sequences that can be directly translated, according to the standard genetic code, to provide an amino acid sequence identical to that translated from a reference nucleic acid sequence.

[0287] The term "gene" or "genomic sequence" as used herein is an open reading frame encoding specific proteins and polypeptides, for example, an mRNA, cDNA, or genomic DNA, and also may or may not include intervening introns, or adjacent 5' and 3' non-coding nucleotide sequences involved in the regulation of expression up to about 20 kb beyond the coding region, and possibly further in either

direction. A gene can be introduced into an appropriate vector for extrachromosomal maintenance or for integration into a host genome.

[0288] The term "transgene" as used herein is a nucleic acid sequence that is incorporated into a transgenic organism. A "transgene" can contain one or more transcriptional regulatory sequences, and other sequences, such as introns, that may be useful for expressing or secreting the nucleic acid or fusion protein it encodes.

[0289] The term "cDNA" as used herein is intended to include all nucleic acids that share the sequence elements of mature mRNA species, where sequence elements are exons and 3' and 5' non-coding regions. Generally, mRNA species have contiguous exons, the intervening introns having been removed by nuclear RNA splicing to create a continuous open reading frame encoding a protein.

[0290] The term "splice variant" refers to all types of RNAs transcribed from a given gene that when processed collectively encode plural protein isoforms. The term "alternative splicing" and related terms refer to all types of RNA processing that lead to expression of plural protein isoforms from a single gene. Some genes are first transcribed as long mRNA precursors that are then shortened by a series of processing steps to produce the mature mRNA molecule. One of these steps is RNA splicing, in which the intron sequences are removed from the mRNA precursor. A cell can splice the primary transcript in different ways, making different "splice variants," and thereby making different polypeptide chains from the same gene, or from the same mRNA molecule. Splice variants can include, for example, exon insertions, exon extensions, exon truncations, exon deletions, alternatives in the 5' untranslated region and alternatives in the 3' untranslated region.

[0291] "Oligonucleotide" may generally refer to polynucleotides of between about 5 and about 100 nucleotides of single-or double-stranded nucleic acids. For the purposes of this disclosure, there is no upper limit to the length of an oligonucleotide. Oligonucleotides are also known as oligomers or oligos and can be isolated from genes, or chemically synthesized by methods known in the art.

[0292] "Nucleic acid composition" as used herein is a composition comprising a nucleic acid sequence, including one having an open reading frame that encodes a polypeptide and is capable, under appropriate conditions, of being expressed as a polypeptide. The term includes, for example, vectors, including plasmids, cosmids, viral vectors (e.g., retrovirus vectors such as lentivirus, adenovirus, and the like), human, yeast, bacterial, P1-derived artificial chromosomes

(HAC's, YAC's, BAC's, PAC's, etc), and mini-chromosomes, *in vitro* host cells, *in vivo* host cells, tissues, organs, allogenic or congenic grafts or transplants, multicellular organisms, and chimeric, genetically modified, or transgenic animals comprising a subject nucleic acid sequence.

[0293] An "isolated," "purified," or "substantially isolated" polynucleotide, or a polynucleotide in "substantially pure form," in "substantially purified form," in "substantial purity," or as an "isolate," is one that is substantially free of the sequences with which it is associated in nature, or other nucleic acid sequences that do not include a sequence or fragment of the subject polynucleotides. By substantially free is meant that less than about 90%, less than about 80%, less than about 70%, less than about 60%, or less than about 50% of the composition is made up of materials other than the isolated polynucleotide. For example, the isolated polynucleotide is at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97%, or at least about 99% free of the materials with which it is associated in nature. For example, an isolated polynucleotide may be present in a composition wherein at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97%, at least about 99% of the total macromolecules (for example, polypeptides, fragments thereof, polynucleotides, fragments thereof, lipids, polysaccharides, and oligosaccharides) in the composition is the isolated polynucleotide. Where at least about 99% of the total macromolecules is the isolated polynucleotide, the polynucleotide is at least about 99% pure, and the composition comprises less than about 1% contaminant. As used herein, an "isolated," "purified" or "substantially isolated" polynucleotide, or a polynucleotide in "substantially pure form," in "substantially purified form," in "substantial purity," or as an "isolate," also refers to recombinant polynucleotides, modified, degenerate and homologous polynucleotides, and chemically synthesized polynucleotides, which, by virtue of origin or manipulation, are not associated with all or a portion of a polynucleotide with which it is associated in nature, are linked to a polynucleotide other than that to which it is linked in nature, or do not occur in nature. For example, the subject polynucleotides are generally provided as other than on an intact chromosome, and recombinant embodiments are typically flanked by one or more nucleotides not normally associated with the subject polynucleotide on a naturally-occurring chromosome.

[0294] The terms "polypeptide," "peptide," and "protein," used interchangeably herein, refer to a polymeric form of amino acids of any length, which can include naturally-occurring amino acids, coded and non-coded amino acids, chemically or biochemically modified, derivatized, or designer amino acids, amino acid analogs, peptidomimetics, and depsipeptides, and polypeptides having modified, cyclic, bicyclic, depsicyclic, or depsibicyclic peptide backbones. The term includes single chain protein as well as multimers. The term also includes conjugated proteins, fusion proteins, including, but not limited to, GST fusion proteins, fusion proteins with a heterologous amino acid sequence, fusion proteins with heterologous and homologous leader sequences, fusion proteins with or without N-terminal methionine residues, pegolyated proteins, and immunologically tagged proteins. Also included in this term are variations of naturally occurring proteins, where such variations are homologous or substantially similar to the naturally occurring protein, as well as corresponding homologs from different species. Variants of polypeptide sequences include insertions, additions, deletions, or substitutions compared with the subject polypeptides. The term also includes peptide aptamers.

[0295] The novel polypeptides herein include amino acid sequences encoded by an open reading frame (ORF) as shown in SEQ ID NOS.: 210 - 418, described in greater detail below, including the full length protein and fragments thereof, particularly biologically active fragments and/or fragments corresponding to functional domains, e.g., a signal peptide or leader sequence, an enzyme active site, including a cleavage site and an enzyme catalytic site, a domain for interaction with other protein(s), a domain for binding DNA, a regulatory domain, a consensus domain that is shared with other members of the same protein family, such as a kinase family or an immunoglobulin family; an extracellular domain that may act as a target for antibody production or that may be cleaved to become a soluble receptor or a ligand for a receptor; an intracellular fragment of a transmembrane protein that participates in signal transduction; a transmembrane domain of a transmembrane protein that may facilitate water or ion transport; a sequence associated with cell survival and/or cell proliferation; a sequence associated with cell cycle arrest, DNA repair and/or apoptosis; a sequence associated with a disease or disease prognosis, including types of cancer, degenerative disease, inflammatory disease, immunological disease, genetic disease, metabolic disease, and/or bacterial or viral infection; and including fusions of the subject polypeptides to other proteins or parts thereof; modifications of the subject

polypeptide, e.g., comprising modified, derivatized, or designer amino acids, modified peptide backbones, and/or immunological tags; as well as intra- and inter-species homologs of the subject polypeptides.

[0296] The term "bicyclic" refers to a peptide with two ring closures formed by covalent linkages between amino acids. A covalent linkage between two nonadjacent amino acids constitutes a ring closure, as does a second covalent linkage between a pair of adjacent amino acids which are already linked by a covalent peptide linkage. The covalent linkages forming the ring closures can be amide linkages, i.e., the linkage formed between a free amino on one amino acid and a free carboxyl of a second amino acid, or linkages formed between the side chains or "R" groups of amino acids in the peptides. Thus, bicyclic peptides can be "true" bicyclic peptides, i.e., peptides cyclized by the formation of a peptide bond between the N-terminus and the C-terminus of the peptide, or they can be "depsi-bicyclic" peptides, i.e., peptides in which the terminal amino acids are covalently linked through their side chain moieties.

[0297] As noted above, a "biologically active" entity, or an entity having "biological activity," is one having structural, regulatory, or biochemical functions of a naturally occurring molecule or any function related to or associated with a metabolic or physiological process. Biologically active polypeptide fragments are those exhibiting activity similar, but not necessarily identical, to an activity of a polypeptide of the present invention. The biological activity can include an improved desired activity, or a decreased undesirable activity. For example, an entity demonstrates biological activity when it participates in a molecular interaction with another molecule, or when it has therapeutic value in alleviating a disease condition, or when it has prophylactic value in inducing an immune response to the molecule, or when it has diagnostic value in determining the presence of the molecule. A biologically active polypeptide or fragment thereof includes one that can participate in a biological reaction, for example, as a transcription factor that combines with other transcription factors for initiation of transcription, or that can serve as an epitope or immunogen to stimulate an immune response, such as production of antibodies, or that can transport molecules into or out of cells, or that can perform a catalytic activity, for example polymerization or nuclease activity, or that can participate in signal transduction by binding to receptors, proteins, or nucleic acids, activating enzymes or substrates.

[0298] A "signal peptide," or a "leader sequence," comprises a sequence of amino acid residues, typically, at the N terminus of a polypeptide, which directs the intracellular trafficking of the polypeptide. Polypeptides that contain a signal peptide or leader sequence typically also contain a signal peptide or leader sequence cleavage site. Such polypeptides, after cleavage at the cleavage sites, generate mature polypeptides, for example, after extracellular secretion or after being directed to the appropriate intracellular compartment.

[0299] "Depsipeptides" are compounds containing a sequence of at least two alpha-amino acids and at least one alpha-hydroxy carboxylic acid, which are bound through at least one normal peptide link and ester links, derived from the hydroxy carboxylic acids. "Linear depsipeptides" can comprise rings formed through S–S bridges, or through an hydroxy or a mercapto group of an hydroxy-, or mercapto-amino acid and the carboxyl group of another amino- or hydroxy-acid but do not comprise rings formed only through peptide or ester links derived from hydroxy carboxylic acids. "Cyclic depsipeptides" are peptides containing at least one ring formed only through peptide or ester links, derived from hydroxy carboxylic acids.

[0300] An "isolated," "purified," or "substantially isolated" polypeptide, or a polypeptide in "substantially pure form," in "substantially purified form," in "substantial purity," or as an "isolate," is one that is substantially free of the materials with which it is associated in nature or other polypeptide sequences that do not include a sequence or fragment of the subject polypeptides. By substantially free is meant that less than about 90%, less than about 80%, less than about 70%, less than about 60%, or less than about 50% of the composition is made up of materials other than the isolated polypeptide. For example, the isolated polypeptide is at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97%, or at least about 99% free of the materials with which it is associated in nature. For example, an isolated polypeptide may be present in a composition wherein at least about 50%, at least about 60%, at least about 70%, at least about 80%, at least about 90%, at least about 95%, at least about 97%, or at least about 99% of the total macromolecules (for example, polypeptides, fragments thereof, polynucleotides, fragments thereof, lipids, polysaccharides, and oligosaccharides) in the composition is the isolated polypeptide. Where at least about 99% of the total macromolecules is the isolated polypeptide, the polypeptide is at least about 99% pure, and the composition comprises less than about 1% contaminant. As

used herein, an "isolated," "purified," or "substantially isolated" polypeptide, or a polypeptide in "substantially pure form," in "substantially purified form," in "substantial purity," or as an "isolate," also refers to recombinant polypeptides, modified, tagged and fusion polypeptides, and chemically synthesized polypeptides, which by virtue or origin or manipulation, are not associated with all or a portion of the materials with which they are associated in nature, are linked to molecules other than that to which they are linked in nature, or do not occur in nature.

[0301] Detection methods of the invention can be qualitative or quantitative. Thus, as used herein, the terms "detection," "identification," "determination," and the like, refer to both qualitative and quantitative determinations, and include "measuring." For example, detection methods include methods for detecting the presence and/or level of polynucleotide or polypeptide in a biological sample, and methods for detecting the presence and/or level of biological activity of polynucleotide or polypeptide in a sample.

[0302] As used herein, the term "array" or "microarray" may be used interchangeably and refers to a collection of plural biological molecules such as nucleic acids, polypeptides, or antibodies, having locatable addresses that may be separately detectable. Generally, "microarray" encompasses use of sub microgram quantities of biological molecules. The biological molecules may be affixed to a substrate or may be in solution or suspension. The substrate can be porous or solid, planar or non-planar, unitary or distributed, such as a glass slide, a 96 well plate, with or without the use of microbeads or nanobeads. As such, the term "microarray" includes all of the devices referred to as microarrays in Schena, 1999; Bassett et al., 1999; Bowtell, 1999; Brown and Botstein, 1999; Chakravarti, 1999; Cheung et al., 1999; Cole et al., 1999; Collins, 1999; Debouck and Goodfellow, 1999; Duggan et al., 1999; Hacia, 1999; Lander, 1999; Lipshutz et al., 1999; Southern, et al., 1999; Schena, 2000; Brenner et al, 2000; Lander, 2001; Steinhaur et al., 2002; and Espejo et al, 2002. Nucleic acid microarrays include both oligonucleotide arrays (DNA chips) containing expressed sequence tags ("ESTs") and arrays of larger DNA sequences representing a plurality of genes bound to the substrate, either one of which can be used for hybridization studies. Protein and antibody microarrays include arrays of polypeptides or proteins, including but not limited to, polypeptides or proteins obtained by purification, fusion proteins, and antibodies, and can be used for specific binding studies (Zhu and Snyder, 2003; Houseman et al., 2002; Schaeferling et al.,

2002; Weng et al., 2002; Winssinger et al., 2002; Zhu et al., 2001; Zhu et al. 2001; and MacBeath and Schreiber, 2000).

[0303] A "nucleic acid hybridization reaction" is one in which single strands of DNA or RNA randomly collide with one another, and bind to each other only when their nucleotide sequences have some degree of complementarity. The solvent and temperature conditions can be varied in the reactions to modulate the extent to which the molecules can bind to one another. Hybridization reactions can be performed under different conditions of "stringency." The "stringency" of a hybridization reaction as used herein refers to the conditions (e.g., solvent and temperature conditions) under which two nucleic acid strands will either pair or fail to pair to form a "hybrid" helix.

[0304] "T_m" is the temperature in degrees Celsius at which 50% of a polynucleotide duplex made of complementary strands of nucleic acids that are hydrogen bonded in an anti-parallel direction by Watson-Crick base pairing dissociate into single strands under conditions of the hybridization reaction. T_m can be predicted according to a standard formula, such as: $T_m = 81.5 + 16.6 \log[X^+] + 0.41$ (%G/C) - 0.61 (%F) - 600/L, where $[X^+]$ is the cation concentration (usually sodium ion, Na^+) in mol/L; (%G/C) is the number of G and C residues as a percentage of total residues in the duplex; (%F) is the percent formamide in solution (wt/vol); and L is the number of nucleotides in each strand of the paired nucleic acids.

[0305] A "buffer" is a system that tends to resist change in pH when a given increment of hydrogen ion or hydroxide ion is added. Buffered solutions contain conjugate acid-base pairs. Any conventional buffer can be used with the inventions herein including but not limited to, for example, Tris, phosphate, imidazole, and bicarbonate.

[0306] A "library" of polynucleotides comprises a collection of sequence information of a plurality of polynucleotide sequences, which information is provided in either biochemical form (e.g., as a collection of polynucleotide molecules), or in electronic form (e.g., as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer-based system, a computer data file, and/or as part of a computer program).

[0307] A "library" of polypeptides comprises a collection of sequence information of a plurality of polypeptide sequences, which information is provided in,

e.g., a collection of polypeptide sequences stored in a computer-readable form, as in a computer-based system, a computer data file, and/or as part of a computer program.

[0308] "Media" refers to a manufacture, other than an isolated nucleic acid molecule, that contains the sequence information of the present invention. Such a manufacture provides the genome sequence or a subset thereof in a form that can be examined by means not directly applicable to the sequence as it exists in a nucleic acid, e.g., with computer-readable media comprising data storage structures. Such media include, but are not limited to: magnetic storage media, such as a floppy disc, a hard disc storage medium, and a magnetic tape; optical storage media such as CD-ROM; electrical storage media such as RAM and ROM; and hybrids of these categories such as magnetic/optical storage media.

[0309] "Recorded" refers to a process for storing information on computer readable media, using any such methods as known in the art.

[0310] As used herein, "a computer-based system" refers to the hardware means, software means, and data storage means used to analyze the nucleotide sequence information of the present invention. The minimum hardware of the computer-based systems of the present invention comprises a central processing unit (CPU), input means, output means, and data storage means. A skilled artisan can readily appreciate that any one of the currently available computer-based systems are suitable for use in the present invention. The data storage means can comprise any manufacture comprising a recording of the present sequence information as described above, or a memory access means that can access such a manufacture.

[0311] "Search means" refers to one or more programs implemented on the computer-based system, to compare a target sequence or target structural motif, or expression levels of a polynucleotide in a sample, with the stored sequence information. A variety of known algorithms are publicly known and commercially available, e.g., MacPattern (EMBL), BLAST, BLASTN and BLASTX (NCBI), gapped BLAST, BLAZE, the Wise package, FASTX, Clustalw, FASTA, FASTA3, Align0, TCoffee, BestFit, FastDB, and TeraBLAST (TimeLogic, Crystal Bay, Nevada). Search means can be used to identify fragments or regions of the genome that match a particular target sequence or target motif, for example, based on sequence similarity, for example, to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms.

[0312] "Sequence similarity," "sequence homology," "homology," "sequence identity," and "percent sequence identity," used interchangeably herein, describe the degree of relatedness between two polynucleotide or polypeptide sequences. In general, "identity" means the exact match-up of two or more nucleotide sequences or two or more amino acid sequences, where the nucleotide or amino acids being compared are the same. Also, in general, "similarity" or "homology" means the exact match-up of two or more nucleotide sequences or two or more amino acid sequences, where the nucleotide or amino acids being compared are either the same or possess similar chemical and/or physical properties. The terms also refer to the percentage of the "aligned" bases (for the polynucleotides) or amino acid residues (for the polypeptides) that are identical when the sequences are aligned. Sequences can be aligned in a number of different ways and sequence similarity can be determined in a number of different ways. For example, the bases or amino acid residues of one sequence can be aligned to a gap in the other sequence, or they can be aligned only to another base or amino acid residue in the other sequence. A gap can range anywhere from one nucleotide, base, or amino acid residue to multiple exons in length, up to any number of nucleotides or amino acid residues. Further, sequences can be aligned such that nucleotides (or bases) align with nucleotides, nucleotides align with amino acid residues, or amino acid residues align with amino acid residues.

[0313] A "target sequence" can be any polynucleotide or amino acid sequence of six or more contiguous nucleotides or two or more amino acids, for example, from about 5 or from about 10 to about 100 amino acids, or from about 15 or from about 30 to about 300 nucleotides. A variety of comparing means can be used to accomplish comparison of sequence information from a sample (e.g., to analyze target sequences, target motifs, or relative expression levels) with the data storage means. A skilled artisan can readily recognize that any one of the publicly available homology search programs can be used as the search means for the computer based systems of the present invention to accomplish comparison of target sequences and motifs. Computer programs to analyze expression levels in a sample and in controls are also known in the art. A "target sequence" includes an "antibody target sequence," which refers to an amino acid sequence that can be used as an immunogen for injection into animals for production of antibodies or for screening against a phage display or antibody library for identification of binding partners.

[0314] A "target structural motif," or "target motif," refers to any rationally selected sequence or combination of sequences in which the sequence(s) are chosen based on a three-dimensional configuration that is formed upon the folding of the target motif, or on consensus sequences of regulatory or active sites. There are a variety of target motifs known in the art. Protein target motifs include, but are not limited to, enzyme active sites and signal sequences. Nucleic acid target motifs include, but are not limited to, hairpin structures, promoter sequences, and other expression elements such as binding sites for transcription factors.

[0315] The term "host cell" includes an individual cell, cell line, cell culture, or *in vivo* cell, which can be or has been a recipient of any polynucleotides or polypeptides of the invention, for example, a recombinant vector, an isolated polynucleotide, antibody or fusion protein. Host cells include progeny of a single host cell, and the progeny may not necessarily be completely identical (in morphology, physiology, or in total DNA, RNA, or polypeptide complement) to the original parent cell due to natural, accidental, or deliberate mutation and/or change. Host cells can be prokaryotic or eukaryotic, including mammalian, insect, amphibian, reptile, crustacean, avian, fish, plant and fungal cells. A host cell includes cells transformed, transfected, transduced, or infected *in vivo* or *in vitro* with a polynucleotide of the invention, for example, a recombinant vector. A host cell which comprises a recombinant vector of the invention may be called a "recombinant host cell."

[0316] The term "agonist" refers to a substance that mimics the function of an active molecule. Agonists include, but are not limited to, drugs, hormones, antibodies, and neurotransmitters, as well as analogues and fragments thereof.

[0317] The term "antagonist" refers to a molecule that competes for the binding sites of an agonist, but does not induce an active response. Antagonists include, but are not limited to, drugs, hormones, antibodies, and neurotransmitters, as well as analogues and fragments thereof.

[0318] The term "receptor" refers to a polypeptide that binds to a specific extracellular molecule and may initiate a cellular response.

[0319] The term "ligand" refers to any molecule that binds to a specific site on another molecule.

[0320] The term "over-expressed" refers to a state wherein there exists any measurable increase over normal or baseline levels. For example, a molecule that is

over-expressed in a disorder is one that is manifest in a measurably higher level compared to levels in the absence of the disorder.

Compositions

[0321] The present invention provides novel isolated polynucleotides encoding polypeptides and fragments thereof. The present invention also provides novel isolated polypeptides, fragments thereof, and compositions comprising same. The present invention further provides polynucleotide compositions that can be used to identify the polypeptides.

[0322] The present invention provides recombinant vectors and host cells for use in gene expression, primer pairs for use in hybridizations, computer-based embodiments for use in bioinformatics, and transgenic animals and embryonic stem cell lines for use in mutating and regulating gene expression.

Nucleic Acids

Sequences

[0323] This invention provides genes encoding proteins, the encoded proteins, and fragments and homologs thereof. It provides human polynucleotide sequences and the corresponding mouse polynucleotide sequences.

[0324] The nucleic acids of the subject invention can encode all or a part of the subject proteins. Double or single stranded fragments can be obtained from the DNA sequence by chemically synthesizing oligonucleotides in accordance with conventional methods, for example by restriction enzyme digestion or polymerase chain reaction (PCR) amplification. The use of the polymerase chain reaction has been described (Saiki et al., 1988) and current techniques have been reviewed (Sambrook et al., 1989; McPherson et al. 2000; Dieffenbach and Dveksler, 1995). For the most part, DNA fragments will be of at least about 5 nucleotides, at least about 8 nucleotides, at least about 10 nucleotides, at least about 15 nucleotides, at least about 18 nucleotides, at least about 20 nucleotides, at least about 25 nucleotides, at least about 30 nucleotides, or at least about 50 nucleotides, at least about 75 nucleotides, or at least about 100 nucleotides. Nucleic acid compositions that encode at least six contiguous amino acids (i.e., fragments of 18 nucleotides or more), for example, nucleic acid compositions encoding at least 8 contiguous amino acids (i.e., fragments of 24 nucleotides or more), are useful in directing the expression or the synthesis of peptides that can be used as immunogens (Lerner, 1982; Shinnick et al., 1983; Sutcliffe et al., 1983).

[0325] In some embodiments, a polynucleotide of the invention comprises a nucleotide sequence of at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, at least about 900, at least about 950, at least about 1000, at least about 1100, at least about 1200, at least about 1300, at least about 1400, at least about 1500, at least about 1600, at least about 1700, at least about 1800, at least about 1900, at least about 2000, at least about 2100, at least about 2200, at least about 2300, at least about 2400, at least about 2500, at least about 3000, at least about 4000, or at least about 5000 contiguous nucleotides of any one of the sequences shown in SEQ ID NOS.: 1 - 209 and 419 - 627, or the coding region thereof, or a complement thereof.

[0326] In other embodiments, a polynucleotide of the invention has at least about 60%, 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 97%, at least about 98%, or at least about 99% nucleotide sequence identity with a nucleotide sequence, or a fragment thereof, of the coding region of any one of the sequences shown in SEQ ID NOS.: 1 - 209 and 419 - 627, or a complement thereof. These sequence variants include naturally-occurring variants (e.g., SNPs, allelic variants, and homologs from other species), degenerate variants, variants associated with disease or pathological states, and variants resulting from random or directed mutagenesis, as well as from chemical or other modification.

[0327] In some embodiments, a polynucleotide of the invention comprises a nucleotide sequence that encodes a polypeptide comprising an amino acid sequence of at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, at least about 900, at least about 950, or at least about 1000 contiguous amino acids of at least one of the sequences shown in SEQ ID NOS.: 210 - 418 (e.g., a polypeptide encoded by at least one of the

nucleotide sequences shown in SEQ ID NOS.: 1 - 209 and 419 - 627), up to and including an entire amino acid sequence as shown in SEQ ID NOS.: 210 - 418 (or as encoded by at least one of the nucleotide sequences shown in SEQ ID NOS.: 1 - 209 and 419 - 627).

[0328] In some embodiment, the present invention includes the present polynucleotide selected from SEQ ID NOS.: 1 - 209 and 419 - 627, which contain 300 bp of 5' terminus of a protein encoding polynucleotide sequence. Such a polynucleotide is useful for the purposes of clustering gene sequences to determine gene family.

[0329] In further embodiments, a polynucleotide of the invention hybridizes under stringent hybridization conditions to a polynucleotide having the coding region of any one of the sequences shown in SEQ ID NOS.: 1 - 209 and 419 - 627, or a complement thereof.

[0330] The polynucleotides of the invention include those that encode variants of the polypeptide sequences encoded by the polynucleotides of the Sequence Listing. In some embodiments, these polynucleotides encode variant polypeptides that include insertions, additions, deletions, or substitutions compared with the polypeptides encoded by the nucleotide sequences shown in SEQ ID NOS.: 1 - 209 and 419 - 627, and in Table 1. Conservative amino acid substitutions include serine/threonine, valine/leucine/isoleucine, asparagine/histidine/glutamine, glutamic acid/aspartic acid, etc. (Gonnet et al., 1992).

[0331] The nucleic acids of the invention include degenerate variants that can be translated, according to the standard genetic code, to provide an amino acid sequence identical to that translated from the nucleic acid sequences herein. For example, synonymous codons include GGG, GGA, GGC, and GGU, each encoding Glycine.

[0332] The nucleic acids of the invention include single nucleotide polymorphisms (SNPs), which occur frequently in eukaryotic genomes (Lander, et al. 2001). The nucleotide sequence determined from one individual of a species can differ from other allelic forms present within the population.

[0333] The nucleic acids of the invention include homologs of the polynucleotides. The source of homologous genes can be any species, e.g., primate species, particularly human; rodents, such as rats, hamsters, guinea pigs, and mice; rabbits, canines, felines; cattles, such as bovines, goats, pigs, sheep, equines,

crustaceans, birds, chickens, reptiles, amphibians, fish, insects, plants, fungi, yeast, nematodes, etc. Among mammalian species, e.g., human and mouse, homologs have substantial sequence similarity, e.g., at least about 60% sequence identity, at least about 75% sequence identity, or at least about 80% sequence identity among nucleotide sequences. In many embodiments of interest, homology will be at least about 75%, at least about 80% ,at least about 85%, at least about 90%, at least about 95%, at least about 97%, or at least about 98%, where in certain embodiments of interest homology will be as high as about 99%.

[0334] Modifications in the native structure of nucleic acids, including alterations in the backbone, sugars or heterocyclic bases, have been shown to increase intracellular stability and binding affinity. Among useful changes in the backbone chemistry are phosphorothioates; phosphorodithioates, where both of the non-bridging oxygens are substituted with sulfur; phosphoroamidites; alkyl phosphotriesters and boranophosphates. Achiral phosphate derivatives include 3'-O'-5'-S-phosphorothioate, 3'-S-5'-O- phosphorothioate, 3'-CH_2-5'-O-phosphonate and 3'-NH-5'-O-phosphoroamidate. Peptide nucleic acids replace the entire ribose phosphodiester backbone with a peptide linkage.

[0335] Sugar modifications are also used to enhance stability and affinity. The α-anomer of deoxyribose can be used, where the base is inverted with respect to the natural β-anomer. The 2'-OH of the ribose sugar can be altered to form 2'-O-methyl or 2'-O-allyl sugars, which provides resistance to degradation without comprising affinity.

[0336] Modification of the heterocyclic bases must maintain proper base pairing. Some useful substitutions include deoxyuridine for deoxythymidine; 5-methyl-2'- deoxycytidine and 5-bromo-2'-deoxycytidine for deoxycytidine. 5-propynyl-2'- deoxyuridine and 5-propynyl-2'-deoxycytidine have been shown to increase affinity and biological activity when substituted for deoxythymidine and deoxycytidine, respectively.

[0337] A genomic sequence of interest comprises the nucleic acid present between the initiation codon and the stop codon, as defined in the listed sequences, including all of the introns that are normally present in a native chromosome. It can further include the 3' and 5' untranslated regions found in the mature mRNA. It can further include specific transcriptional and translational regulatory sequences, such as promoters, enhancers, etc., including about 1 kb, about 2 kb, and possibly more, of

flanking genomic DNA at either the 5' or 3' end of the transcribed region. The genomic DNA can be isolated as a fragment of 100 kbp or smaller; and substantially free of flanking chromosomal sequence. The genomic DNA flanking the coding region, either 3' or 5', or internal regulatory sequences as sometimes found in introns, contains sequences required for proper tissue and stage specific expression.

[0338] Nucleic acid molecules of the invention can comprise heterologous nucleic acid molecules, i.e., nucleic acid molecules other than the subject nucleic acid molecules, of any length. For example, the subject nucleic acid molecules can be flanked on the 5' and/or 3' ends by heterologous nucleic acid molecules of from about 1 nucleotide to about 10 nucleotides, from about 10 nucleotides to about 20 nucleotides, from about 20 nucleotides to about 50 nucleotides, from about 50 nucleotides to about 100 nucleotides, from about 100 nucleotides to about 250 nucleotides, from about 250 nucleotides to about 500 nucleotides, or from about 500 nucleotides to about 1000 nucleotides, or more in length.

[0339] The subject polynucleotides include those that encode fusion proteins comprising the subject polypeptides fused to "fusion partners." For example, the present soluble receptor or ligand can be fused to an immunoglobulin fragment, such as an Fc fragment for stability in circulation or to fix complement. Other polypeptide fragments that have equivalent capabilities as the Fc fragments can also be used herein.

[0340] The isolated nucleic acids of the invention can be used as probes to detect and characterize gross alteration in a genomic locus, such as deletions, insertions, translocations, and duplications, e.g., applying fluorescence *in situ* hybridization (FISH) techniques to examine chromosome spreads (Andreeff et al., 1999). The nucleic acids are also useful for detecting smaller genomic alterations, such as deletions, insertions, additions, translocations, and substitutions (e.g., SNPs).

[0341] When used as probes to detect nucleic acid molecules capable of hybridizing with nucleic acids described in the Sequence Listing, the nucleic acid molecules can be flanked by heterologous sequences of any length. When used as probes, a subject nucleic acid can include nucleotide analogs that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogs that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin, digoxigenin, and dinitrophenyl.

[0342] Suitable fluorescent labels include fluorochromes e.g., fluorescein and its derivatives, e.g., fluorescein isothiocyanate (FITC6-carboxyfluorescein (6-FAM), 2',7'-dimethoxy-4',5'-dichloro-6-carboxyfluorescein (JOE),), 6-carboxy-2',4',7',4,7-hexachlorofluorescein (HEX), 5-carboxyfluorescein (5-FAM); coumarin and its derivatives, e.g., 7-amino-4-methylcoumarin, aminocoumarin; bodipy dyes, such as Bodipy FL; cascade blue; Oregon green; rhodamine dyes, e.g., rhodamine, 6-carboxy-X-rhodamine (ROX), Texas red, phycoerythrin, and tetramethylrhodamine; eosins and erythrosins; cyanine dyes, e.g., allophycocyanin, Cy3 and Cy5 or N,N,N',N'-tetramethyl-6-carboxyrhodamine (TAMRA); macrocyclic chelates of lanthanide ions, e.g., quantum dye, etc; and chemiluminescent molecules, e.g., luciferases.

[0343] Fluorescent labels also include a green fluorescent protein (GFP), i.e., a "humanized" version of a GFP, e.g., wherein codons of the naturally-occurring nucleotide sequence are changed to more closely match human codon bias; a GFP derived from *Aequoria victoria* or a derivative thereof, e.g., a "humanized" derivative such as Enhanced GFP, which are available commercially, e.g., from Clontech, Inc.; other fluorescent mutants of a GFP from *Aequoria victoria*, e.g., as described in U.S. Patent No. 6,066,476; 6,020,192; 5,985,577; 5,976,796; 5,968,750; 5,968,738; 5,958,713; 5,919,445; 5,874,304; a GFP from another species such as *Renilla reniformis, Renilla mulleri,* or *Ptilosarcus guernyi*, as previously described (WO 99/49019; Peelle et al., 2001), "humanized" recombinant GFP (hrGFP) (Stratagene®); any of a variety of fluorescent and colored proteins from Anthozoan species, (e.g., Matz et al., 1999).

[0344] Probes can also contain fluorescent analogs, including commercially available fluorescent nucleotide analogs that can readily be incorporated into a subject nucleic acid. These include deoxyribonucleotides and/or ribonucleotide analogs labeled with Cy3, Cy5, Texas Red, Alexa Fluor dyes, rhodamine, cascade blue, or BODIPY, and the like.

[0345] Suitable radioactive labels include, e.g., ^{32}P, ^{35}S, or ^{3}H. For example, probes can contain radiolabeled analogs, including those commonly labeled with ^{32}P or ^{35}S, such as α-^{32}P-dATP, -dTTP, -dCTP, and dGTP; γ-^{35}S-GTP and α-^{35}S-dATP, and the like.

[0346] Nucleic acids of the invention can also be bound to a substrate. Subject nucleic acids can be attached covalently, attached to a surface of the support

or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence, e.g., by noncovalent interactions, or some combination thereof. The nucleic acids can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of the bound nucleic acids being separately detectable.

[0347] The substrate can be porous or solid, planar or non-planar, unitary or distributed; and the bond between the nucleic acid and the substrate can be covalent or non-covalent. The substrate can be in the form of microbeads or nanobeads. Substrates include, but are not limited to, a membrane, such as nitrocellulose, nylon, positively-charged derivatized nylon; a solid substrate such as glass, amorphous silicon, crystalline silicon, plastics (including e.g., polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, cellulose acetate, or mixtures thereof).

[0348] The subject nucleic acids include antisense RNA, ribozymes, and RNAi. Further, The nucleic acids of the invention can be used for antisense or RNAi inhibition of transcription or translation using methods known in the art (Phillips, 1999a; Phillips, 1999b; Hartmann et al., 1999; Stein et al., 1998; Agrawal et al., 1998).

Expression Vectors

[0349] The instant invention further provides host cells, e.g., recombinant host cells, that comprise a subject nucleic acid, host cells that comprise a recombinant vector, and host cells that secrete antibodies of the invention. Subject host cells can be cultured *in vitro*, or can be part of a multicellular organism. Host cells are described in more detail below. The instant invention further provides transgenic plants and non-human animals, as described in more detail below.

[0350] In addition to the plurality of uses described in greater detail in following sections, the subject nucleic acids find use in the preparation of all or a portion of the polypeptides of the subject invention, as described above, using an expression system. For expression, an expression vector can be employed. The expression vector will provide a transcriptional and translational initiation region, which may be inducible, conditionally-active, or constitutive, or tissue-specific, where the coding region is operably linked under the transcriptional control of the transcriptional initiation region, and a transcriptional and translational termination

region. These control regions can be native to a gene encoding the subject peptides, or can be derived from heterologous or exogenous sources.

[0351] The subject nucleic acids can also be provided as part of a vector (e.g., a polynucleotide construct comprising an expression cassette), a wide variety of which are known in the art. Vectors include, but are not limited to, plasmids; cosmids; viral vectors; human, yeast, bacterial, P1-derived artificial chromosomes (HAC's, YAC's, BAC's, PAC's, etc.), mini-chromosomes, and the like. Vectors are amply described in numerous publications well known to those in the art (Ausubel, et al.; Jones et al., 1998a; Jones et al., 1998b). Vectors can provide for nucleic acid expression, for nucleic acid propagation, or both.

[0352] A recombinant vector or construct that includes a nucleic acid of the invention is useful for propagating a nucleic acid in a host cell; such vectors are known as "cloning vectors." Vectors can transfer nucleic acid between host cells derived from disparate organisms; these are known in the art as "shuttle vectors." Vectors can also insert a subject nucleic acid into a host cell's chromosome; these are known in the art as "insertion vectors." Vectors can express either sense or antisense RNA transcripts of the invention *in vitro* (e.g., in a cell-free system or within an *in vitro* cultured host cell) or *in vivo* (e.g., in a multicellular plant or animal); these are known in the art as "expression vectors," which can be part of an expression system. Expression vectors can also produce a subject antibody.

Vectors typically include at least one origin of replication, at least one site for insertion of heterologous nucleic acid (e.g., in the form of a polylinker with multiple, tightly clustered, single cutting restriction endonuclease recognition sites), and at least one selectable marker, although some integrative vectors will lack an origin that is functional in the host to be chromosomally modified, and some vectors will lack selectable markers. Vectors are transiently or stably be maintained in the cells, usually for a period of at least about one day, at least about several days to at least about several weeks.

[0353] Prior to vector insertion, the DNA of interest will be obtained substantially free of other nucleic acid sequences. The DNA can be "recombinant," and flanked by one or more nucleotides with which it is not normally associated on a naturally occurring chromosome.

[0354] Expression vectors generally have convenient restriction sites located near the promoter sequence to provide for the insertion of nucleic acid sequences

encoding heterologous protein or RNA molecules. A selectable marker operative in the expression system or host can be present. Expression vectors can be used for the production of fusion proteins, where the fusion peptide provides additional functionality, i.e., increased protein synthesis, a leader sequence for secretion, stability, reactivity with defined antisera, or an enzyme marker, e.g., β-galactosidase.

[0355] Promoters of the invention can be naturally contiguous or not naturally contiguous to the expressed nucleic acid molecule. The promoters can be inducible, conditionally active (such as the cre-lox promoter), constitutive, and/or tissue specific.

[0356] Expression vectors can be prepared comprising a transcription cassette comprising a transcription initiation region, the gene or fragment thereof, and a transcriptional termination region. Of particular interest is the use of DNA sequences that allow for the expression of functional epitopes or domains, at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, at least about 900, at least about 950, or at least about 1000 amino acids in length, or any of the above-described fragments, up to and including the complete open reading frame of the gene. After introduction of these DNA sequences, the cells containing the vector construct can be selected by means of a selectable marker, and the selected cells expanded and used as expression-competent host cells.

[0357] Host cells can comprise prokaryotes or eukaryotes that express proteins and polypeptides in accordance with conventional methods, the method depending on the purpose for expression. For large scale production of the protein, a unicellular organism, such as *E. coli, B. subtilis, S. cerevisiae,* insect cells in combination with baculovirus vectors, or cells of a higher organism such as vertebrates, particularly mammals, e.g., COS 7 cells, can be used as the expression host cells. In some situations, it is desirable to express eukaryotic genes in eukaryotic cells, where the encoded protein will benefit from native folding and post-translational modifications.

[0358] Specific expression systems of interest include plants, bacteria, yeast, insect cells, and mammalian cell-derived expression systems. Representative systems from each of these categories are provided below.

[0359] Expression systems in plants include those described in U.S. Patent No. 6,096,546 and U.S. Patent No. 6,127,145.

[0360] Expression systems in bacteria include those described by Chang et al., 1978; Goeddel et al., 1979; Goeddel et al., 1980; EP 0 036,776; U.S. Patent No. 4,551,433; DeBoer et al., 1983); and Siebenlist et al., 1980.

[0361] Expression systems in yeast include those described by Hinnen et al., 1978; Ito et al., 1983; Kurtz et al., 1986; Kunze et al., 1985; Gleeson et al., 1986; Roggenkamp et al., 1986; Das et al., 1984; De Louvencourt et al., 1983; Van den Berg et al., 1990; Kunze et al., 1985; Cregg et al., 1985; U.S. Patent Nos. 4,837,148 and 4,929,555; Beach and Nurse, 1981; Davidow et al., 1985; Gaillardin et al., 1985; Ballance et al., 1983; Tilburn et al., 1983; Yelton et al., 1984; Kelly and Hynes, 1985; EP 0 244,234; WO 91/00357; and U.S. Patent No. 6,080,559.

[0362] Expression systems for heterologous genes in insects include those described in U.S. Patent No. 4,745,051; Friesen et al., 1986; EP 0 127,839; EP 0 155,476; Vlak et al., 1988; Miller et al., 1988; Carbonell et al., 1988; Maeda et al., 1985; Lebacq-Verheyden et al., 1988; Smith et al., 1985); Miyajima et al., 1987; and Martin et al., 1988. Numerous baculoviral strains and variants and corresponding permissive insect host cells are described in Luckow et al., 1988, Miller et al., 1986, and Maeda et al., 1985. The insect cell expression system is useful not only for production of heterologous proteins intracellularly, but can be used for expression of transmembrane proteins on the insect cell surfaces. Such insect cells can be used as immunogen for production of antibodies, for example, by injection of the insect cells into mice or rabbits or other suitable animals, for production of antibodies.

[0363] Mammalian expression systems include those described in Dijkema et al., 1985; Gorman et al., 1982; Boshart et al., 1985; and U.S. Patent No. 4,399,216. Additional features of mammalian expression are facilitated as described in Ham and Wallace, 1979; Barnes and Sato, 1980 U.S. Patent Nos. 4,767,704, 4,657,866, 4,927,762, 4,560,655, WO 90/103430, WO 87/00195, and U.S. RE 30,985. Mammalian cell expression systems can also be used for production of antibodies.

[0364] The present polynucleotides can also be used in cell-free expression systems such as bacterial system, e.g., *E. coli* lysate, rabbit reticulocyte lysate system,

wheat germ extract system, frog oocyte lysate system, and the like which is conventional in the art. See, for example, WO 00/68412, WO 01/27260, WO 02/24939, WO 02/38790, WO 91/02076, and WO 91/02075.

[0365] When any of the above-referenced host cells, or other appropriate host cells or organisms, are used to replicate and/or express the polynucleotides of the invention, the resulting replicated nucleic acid, RNA, expressed protein or polypeptide, is within the scope of the invention as a product of the host cell or organism.

[0366] Once the gene corresponding to a selected polynucleotide is identified, its expression can be regulated in the gene's native cell types. For example, an endogenous gene of a cell can be regulated by an exogenous regulatory sequence inserted into the genome of the cell at a location that will enhance or reduce expression of the gene corresponding to the subject polypeptide. The regulatory sequence can be designed to integrate into the genome via homologous recombination, as disclosed in U.S. Patent Nos. 5,641,670 and 5,733,761, the disclosures of which are herein incorporated by reference. Alternatively, it can be designed to integrate into the genome via non-homologous recombination, as described in WO 99/15650, the disclosure of which is also herein incorporated by reference. Also encompassed in the subject invention is the production of proteins without manipulating the encoding nucleic acid itself, but rather by integrating a regulatory sequence into the genome of a cell that already includes a gene that encodes the protein of interest; this production method is described in the above-incorporated patent documents.

Isolated Primer Pairs

[0367] In some embodiments, the invention provides isolated nucleic acids that, when used as primers in a polymerase chain reaction, amplify a subject polynucleotide, or a polynucleotide containing a subject polynucleotide. The amplified polynucleotide is from about 20 to about 50, from about 50 to about 75, from about 75 to about 100, from about 100 to about 125, from about 125 to about 150, from about 150 to about 175, from about 175 to about 200, from about 200 to about 250, from about 250 to about 300, from about 300 to about 350, from about 350 to about 400, from about 400 to about 500, from about 500 to about 600, from about 600 to about 700, from about 700 to about 800, from about 800 to about 900, from about 900 to about 1000, from about 1000 to about 2000, from about 2000 to about

3000, from about 3000 to about 4000, from about 4000 to about 5000, or from about 5000 to about 6000 nucleotides or more in length.

[0368] The isolated nucleic acids themselves are from about 10 to about 20, from about 20 to about 30, from about 30 to about 40, from about 40 to about 50, from about 50 to about 100, or from about 100 to about 200 nucleotides in length. Generally, the nucleic acids are used in pairs in a polymerase chain reaction, where they are referred to as "forward" and "reverse" primers.

[0369] Thus, in some embodiments, the invention provides a pair of isolated nucleic acid molecules, each from about 10 to about 200 nucleotides in length, the first nucleic acid molecule of the pair comprising a sequence of at least 10 contiguous nucleotides having 100% sequence identity to a nucleic acid sequence as shown in SEQ ID NOS.: 1 - 209 and 419 - 627 and the second nucleic acid molecule of the pair comprising a sequence of at least 10 contiguous nucleotides having 100% sequence identity to the reverse complement of the nucleic acid sequence shown in SEQ ID NOS.: 1 - 209 and 419 - 627 , wherein the sequence of the second nucleic acid molecule is located 3' of the nucleic acid sequence of the first nucleic acid molecule shown in SEQ ID NOS.: 1 - 209 and 419 - 627. The primer nucleic acids are prepared using any known method, e.g., automated synthesis, and can be chosen to specifically amplify a cDNA copy of an mRNA encoding a subject polypeptide.

[0370] In some embodiments, the first and/or the second nucleic acid molecules comprise a detectable label. The label can be a radioactive molecule, fluorescent molecule or another molecule, e.g., hapten, as described in detail above. Further, the label can be a two stage system, where the amplified DNA is conjugated to another molecule, i.e., biotin, digoxin, or a hapten, that has a high affinity binding partner, i.e., avidin, antidigoxin, or a specific antibody, respectively, and the binding partner conjugated to a detectable label. The label can be conjugated to one or both of the primers. Alternatively, the pool of nucleotides used in the amplification is labeled, so as to incorporate the label into the amplification product.

[0371] Conditions that increase stringency of both DNA/DNA and DNA/RNA hybridization reactions are widely known and published in the art. See, for example, Sambrook, 1989, and examples provided above. Examples of relevant conditions include (in order of increasing stringency): incubation temperatures of 25°C, 37°C, 50°C, and 68°C; buffer concentrations of 10 x SSC, 6 x SSC, 1 x SSC, 0.1 x SSC (where 1 x SSC is 0.15 M NaCl and 15 mM citrate buffer); and their

equivalents using other buffer systems; formamide concentrations of 0%, 25%, 50%, and 75%; incubation times from 5 minutes to 24 hours; 1, 2, or more washing steps; wash incubation times of 1, 2, or 15 minutes; and wash solutions of 6 x SSC, 1 x SSC, 0.1 x SSC, or deionized water.

[0372] For example, "high stringency conditions" include hybridization in 50% formamide, 5X SSC, 0.2 μg/μl poly(dA), 0.2 μg/μl human cot1 DNA, and 0.5% SDS, in a humid oven at 42°C overnight, followed by successive washes in 1X SSC, 0.2% SDS at 55°C for 5 minutes, followed by washing at 0.1X SSC, 0.2% SDS at 55°C for 20 minutes. Further examples of high stringency conditions include hybridization at 50°C and 0.1×SSC (15 mM sodium chloride/1.5 mM sodium citrate); overnight incubation at 42°C in a solution containing 50% formamide, 1 × SSC (150 mM NaCl, 15 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5 × Denhardt's solution, 10% dextran sulfate, and 20 μg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1 × SSC at about 65°. High stringency conditions also include aqueous hybridization (e.g., free of formamide) in 6X SSC (where 20X SSC contains 3.0 M NaCl and 0.3 M sodium citrate), 1% sodium dodecyl sulfate (SDS) at 65°C for about 8 hours (or more), followed by one or more washes in 0.2 X SSC, 0.1% SDS at 65°C. Highly stringent hybridization conditions are hybridization conditions that are at least as stringent as any one of the above representative conditions. Other stringent hybridization conditions are known in the art and can also be employed to identify nucleic acids of this particular embodiment of the invention.

[0373] Conditions of "reduced stringency," suitable for hybridization to molecules encoding structurally and functionally related proteins, or otherwise serving related or associated functions, are the same as those for high stringency conditions but with a reduction in temperature for hybridization and washing to lower temperatures (e.g., room temperature or about 22°C to 25°C). For example, moderate stringency conditions include aqueous hybridization (e.g., free of formamide) in 6X SSC, 1% SDS at 65°C for about 8 hours (or more), followed by one or more washes in 2X SSC, 0.1% SDS at room temperature. Low stringency conditions include, for example, aqueous hybridization at 50°C and 6×SSC (0.9 M sodium chloride/0.09 M sodium citrate) and washing at 25°C in 1×SSC (0.15 M sodium chloride/0.015 M sodium citrate).

[0374] The specificity of a hybridization reaction allows any single-stranded sequence of nucleotides to be labeled with a radioisotope or chemical and used as a probe to find a complementary strand, even in a cell or cell extract that contains millions of different DNA and RNA sequences. Probes of this type are widely used to detect the nucleic acids corresponding to specific genes, both to facilitate the purification and characterization of the genes after cell lysis and to localize them in cells, tissues, and organisms.

[0375] Moreover, by carrying out hybridization reactions under conditions of "reduced stringency," a probe prepared from one gene can be used to find homologous evolutionary relatives - both in the same organism, where the relatives form part of a gene family, and in other organisms, where the evolutionary history of the nucleotide sequence can be traced. A person skilled in the art would recognize how to modify the conditions to achieve the requisite degree of stringency for a particular hybridization.

Libraries

[0376] The polynucleotide libraries of the invention generally comprise a collection of sequence information of a plurality of polynucleotide sequences, where at least one of the polynucleotides has a sequence shown in SEQ ID NOS.: 1 -209 and 419 - 627. By plurality is meant at least 2, at least 3, or at least all of the sequences in the Sequence Listing. The information may be provided in either biochemical form (e.g., as a collection of polynucleotide molecules), or in electronic form (e.g., as a collection of polynucleotide sequences stored in a computer-readable form, as in a computer-based system, a computer data file, and/or as a part of a computer program). The length and number of polynucleotides in the library will vary with the nature of the library, e.g., if the library is an oligonucleotide array, a cDNA array, or a computer database of the sequence information.

[0377] The sequence information contained in either a biochemical or an electronic library of polynucleotides can be used in a variety of ways, e.g., as a resource for gene discovery, as a representation of sequences expressed in a selected cell type (e.g., cell type markers), or as markers of a given disorder or disease state. In general, a disease marker is a representation of a gene product that is present in all cells affected by disease either at an increased or decreased level relative to a normal cell (e.g., a cell of the same or similar type that is not substantially affected by disease). For example, a polynucleotide sequence in a library can be a polynucleotide

that represents an mRNA, polypeptide, or other gene product encoded by the polynucleotide, that is either over-expressed or under-expressed in one cell compared to another (e.g., a first cell type compared to a second cell type; a normal cell compared to a diseased cell; a cell not exposed to a signal or stimulus compared to a cell exposed to that signal or stimulus; and the like).

[0378] The nucleotide sequence information of the library can be embodied in any suitable form, e.g., electronic or biochemical forms. For example, a library of sequence information embodied in electronic form comprises an accessible computer data file that may contain the representative nucleotide sequences of genes that are differentially expressed (e.g., over-expressed or under-expressed) as between, e.g., a first cell type compared to a second cell type (e.g., expression in a brain cell compared to expression in a kidney cell); a normal cell compared to a diseased cell (e.g., a non-cancerous cell compared to a cancerous cell); a cell not exposed to an internal or external signal or stimulus compared to a cell exposed to that signal or stimulus (e.g., a cell contacted with a ligand compared to a control cell not contacted with the ligand); and the like. Other combinations and comparisons of cells will be readily apparent to the ordinarily skilled artisan. Biochemical embodiments of the library include a collection of nucleic acid molecules that have the sequences of the genes in the library, where the nucleic acids can correspond to the entire gene in the library or to a fragment thereof, as described in greater detail below.

[0379] Where the library is an electronic library, the nucleic acid sequence information can be present in a variety of media. For example, the nucleic acid sequences of any of the polynucleotides shown in SEQ ID NOS.: 1 -209 and 419 - 627 can be recorded on computer readable media of a computer-based system, e.g., any medium that can be read and accessed directly by a computer. One of skill in the art can readily appreciate how any of the presently known computer readable mediums can be used to create a manufacture comprising a recording of the present sequence information. Any convenient data storage structure can be chosen, based on the means used to access the stored information. A variety of data processor programs and formats can be used for storage, e.g., word processing text file, database format, etc. In addition to the sequence information, electronic versions of the libraries of the invention can be provided in conjunction or connection with other computer-readable information and/or other types of computer-based files (e.g.,

searchable files, executable files, etc, including, but not limited to, for example, search program software, etc.).

[0380] By providing the nucleotide sequence in computer readable form in a computer-based system, the information can be accessed for a variety of purposes. Computer software to access sequence information is publicly available. Conventional bioinformatics tools can be utilized to analyze sequences to determine sequence identity, sequence similarity, and gap information. For example, the gapped BLAST (Altschul et al., 1990, Altschul et al., 1997), and BLAZE (Brutlag et al., 1993) search algorithms on a Sybase system, or the TeraBLAST (TimeLogic, Crystal Bay, Nevada) program optionally running on a specialized computer platform available from TimeLogic, can be used to identify open reading frames (ORFs) within the genome that contain homology to ORFs from other organisms. Homology between sequences of interest can be determined using the local homology algorithm of Smith and Waterman, 1981, as well as the BestFit program (Rechid et al., 1989), and the FastDB algorithm (FastDB, 1988; described in Current Methods in Sequence Comparison and Analysis, Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp. 127-149, 1988, Alan R. Liss, Inc).

[0381] Alignment programs that permit gaps in the sequence include Clustalw (Thompson et al., 1994), FASTA3 (Pearson, 2000) Align0 (Myers and Miller, 1988), and TCoffee (Notredame et al., 2000). Other methods for comparing and aligning nucleotide and protein sequences include, for example, BLASTX (NCBI), the Wise package (Birney and Durbin, 2000), and FASTX (Pearson, 2000). These algorithms determine sequence homology between nucleotide and protein sequences without translating the nucleotide sequences into protein sequences. Other techniques for alignment are also known in the art (Doolittle, et al., 1996; BLAST, available from the National Center for Biotechnology Information; FASTA, available in the Genetics Computing Group (GCG) package, from Madison, Wisconsin, USA, a wholly owned subsidiary of Oxford Molecular Group, Inc.; Schlessinger, 1988a; Schlessinger, 1988b; and Needleman and Wunch, 1970).

[0382] Sequence similarity is calculated based on a reference sequence, which may be a subset of a larger sequence, such as a conserved motif, coding region, flanking region, etc. The reference sequence is usually at least about 18 nt long, at least about 30 nt long, or may extend to the complete sequence that is being compared.

[0383] One parameter for determining percent sequence identity is the percentage of the alignment in the region of strongest alignment between a target and a query sequence. Methods for determining this percentage involve, for example, counting the number of aligned bases of a query sequence in the region of strongest alignment and dividing this number by the total number of bases in the region. For example, 10 matches divided by 11 total residues gives a percent sequence identity of approximately 90.9%. The length of the aligned region is typically at least about 55%, at least about 58%, or at least about 60% of the total sequence length, and can be as great as about 62%, as great as about 64%, and even as great as about 66% of the total sequence length.

[0384] The present invention includes human and mouse polynucleotide and polypeptide sequences that are at least about 95%, at least about 96%, at least about 97%, at least about 98%, or at least about 99% homologous to the sequences in the Sequence Listing, based on using the method of determining sequence identity with the insertion of gaps to detect the maximum degree of sequence identity. In other embodiments of interest, homology will be at least about 80%, at least about 85%, or as high as about 90%.

[0385] A variety of structural formats for the input and output means can be used to input and output the information in the computer-based systems of the present invention. One format for an output means ranks the relative expression levels of different polynucleotides. Such presentation provides a skilled artisan with a ranking of relative expression levels to determine a gene expression profile.

[0386] As discussed above, the library of the invention also encompasses biochemical libraries of the polynucleotides shown in SEQ ID NOS.: 1 - 209 and 419 - 627, e.g., collections of nucleic acids representing the provided polynucleotides. The biochemical libraries can take a variety of forms, e.g., a solution of cDNAs, a pattern of probe nucleic acids stably associated with a surface of a solid support (i.e., an array) and the like. Of particular interest are nucleic acid arrays in which one or more of the polynucleotide sequences shown in SEQ ID NOS.: 1 - 209 and 419 - 627 is represented on the array. A variety of different array formats have been developed and are known to those of skill in the art. The arrays of the subject invention find use in a variety of applications, including gene expression analysis, drug screening, mutation analysis, and the like, as disclosed in the herein-listed exemplary patent documents.

[0387] In addition to the above nucleic acid libraries, analogous libraries of polypeptides are also provided, where the polypeptides of the library will represent at least a portion of the polypeptides encoded by a gene corresponding to one or more of the sequences shown in SEQ ID NOS.: 1 -209 and 419 - 627.

[0388] Further, analogous libraries of antibodies are also provided, where the libraries comprise antibodies or fragments thereof that specifically bind to at least a portion of at least one of the subject polypeptides. Further, antibody libraries may comprise antibodies or fragments thereof that specifically inhibit binding of a subject polypeptide to its ligand or substrate, or that specifically inhibit binding of a subject polypeptide as a substrate to another molecule. Moreover, corresponding nucleic acid libraries are also provided, comprising polynucleotide sequences that encode the antibodies or antibody fragments described above.

Polypeptides

Peptides and Modified Peptides

[0389] In some embodiments of the present invention, the active agent is a peptide. Suitable peptides include peptides of from about 3 amino acids to about 50, from about 5 to about 30, or from about 10 to about 25 amino acids in length. In some embodiments, a peptide has a sequence of from about 3 amino acids to about 50, from about 5 to about 30, or from about 10 to about 25 amino acids of corresponding naturally-occurring protein. In some embodiments, a peptide exhibits one or more of the following activities: inhibits binding of a subject polypeptide to an interacting protein or other molecule; inhibits subject polypeptide binding to a second polypeptide molecule; inhibits a signal transduction activity of a subject polypeptide; inhibits an enzymatic activity of a subject polypeptide; or inhibits a DNA binding activity of a subject polypeptide.

[0390] This invention provides novel polypeptides, and related polypeptide compositions. The novel polypeptides of the invention encompass proteins with amino acid sequences as shown in SEQ ID NOS.: 210 - 418, or encoded by the nucleic acids having nucleotide sequences shown in SEQ ID NOS.: 1 -209 and 419 - 627. The subject polypeptides are human polypeptides, fragments thereof, variants (such as splice variants), homologs from other species, and derivatives thereof. In particular embodiments, a polypeptide of the invention has an amino acid sequence substantially identical to the sequence of any polypeptide encoded by a polynucleotide sequence shown in SEQ ID NOS.: 1 -209 and 419 - 627.

[0391] Peptides can include naturally-occurring and non-naturally occurring amino acids. Peptides can comprise D-amino acids, a combination of D- and L-amino acids, and various "designer" amino acids (e.g., β-methyl amino acids, Cα-methyl amino acids, and Nα-methyl amino acids, etc.) to convey special properties. Additionally, peptides can be cyclic. Peptides can include non-classical amino acids in order to introduce particular conformational motifs. Any known non-classical amino acid can be used. Non-classical amino acids include, but are not limited to, 1,2,3,4-tetrahydroisoquinoline-3-carboxylate; (2S,3S)-methylphenylalanine, (2S,3R)-methyl-phenylalanine, (2R,3S)-methyl-phenylalanine and (2R,3R)-methyl-phenylalanine; 2-aminotetrahydronaphthalene-2-carboxylic acid; hydroxy-1,2,3,4-tetrahydroisoquinoline-3-carboxylate; β-carboline (D and L); HIC (histidine isoquinoline carboxylic acid); and HIC (histidine cyclic urea). Amino acid analogs and peptidomimetics can be incorporated into a peptide to induce or favor specific secondary structures, including, but not limited to, LL-Acp (LL-3-amino-2-propenidone-6-carboxylic acid), a β-turn inducing dipeptide analog; β-sheet inducing analogs; β-turn inducing analogs; α-helix inducing analogs; γ-turn inducing analogs; Gly-Ala turn analogs; amide bond isostere; or tretrazol, and the like.

[0392] A peptide can be a depsipeptide, which can be linear or cyclic (Kuisle et al., 1999). Linear depsipeptides can comprise rings formed through S–S bridges, or through an hydroxy or a mercapto group of an hydroxy-, or mercapto-amino acid and the carboxyl group of another amino- or hydroxy-acid but do not comprise rings formed only through peptide or ester links derived from hydroxy carboxylic acids. Cyclic depsipeptides contain at least one ring formed only through peptide or ester links, derived from hydroxy carboxylic acids.

[0393] Peptides can be cyclic or bicyclic. For example, the C-terminal carboxyl group or a C-terminal ester can be induced to cyclize by internal displacement of the -OH or the ester (-OR) of the carboxyl group or ester respectively with the N-terminal amino group to form a cyclic peptide. For example, after synthesis and cleavage to give the peptide acid, the free acid is converted to an activated ester by an appropriate carboxyl group activator such as dicyclohexylcarbodiimide (DCC) in solution, for example, in methylene chloride (CH_2Cl_2), dimethyl formamide (DMF) mixtures. The cyclic peptide is then formed by internal displacement of the activated ester with the N-terminal amine. Internal

cyclization as opposed to polymerization can be enhanced by use of very dilute solutions. Methods for making cyclic peptides are well known in the art.

[0394] A desamino or descarboxy residue can be incorporated at the terminal ends of the peptide, so that there is no terminal amino or carboxyl group, to decrease susceptibility to proteases or to restrict conformation. C-terminal functional groups include amide, amide lower alkyl, amide di (lower alkyl), lower alkoxy, hydroxy, and carboxy, and the lower ester derivatives thereof, and the pharmaceutically acceptable salts thereof.

[0395] In addition to the foregoing N-terminal and C-terminal modifications, a peptide or peptidomimetic can be modified with or covalently coupled to one or more of a variety of hydrophilic polymers to increase solubility and circulation half-life of the peptide. Suitable nonproteinaceous hydrophilic polymers for coupling to a peptide include, but are not limited to, polyalkylethers as exemplified by polyethylene glycol and polypropylene glycol, polylactic acid, polyglycolic acid, polyoxyalkenes, polyvinylalcohol, polyvinylpyrrolidone, cellulose and cellulose derivatives, dextran, and dextran derivatives. Generally, such hydrophilic polymers have an average molecular weight ranging from about 500 to about 100,000 daltons, from about 2,000 to about 40,000 daltons, or from about 5,000 to about 20,000 daltons. The peptide can be derivatized with or coupled to such polymers using any of the methods set forth in Zallipsky, 1995; Monfardini et al., 1995; U.S. Pat. Nos. 4,640,835; 4,496,689; 4,301,144; 4,670,417; 4,791,192; 4,179,337, or WO 95/34326.

[0396] These polypeptides may reside within the cell, or extracellularly. They may be secreted from the cell, reside in the cytoplasm, in the membranes, or in any of the intracellular organelles, including the nucleus, mitochondria, ribosomes, or storage granules.

[0397] In many embodiments, a novel polypeptide of the invention functions as a secreted protein, a single-transmembrane protein, a multiple-transmembrane protein, a kinase, a protein kinase, a ligase, a nuclear hormone receptor, a phosphatase, a protease, a phosphodiesterase, a kinesin, an immunoglobulin, a T-cell receptor, or a glycosylphosphatidylinositol anchor. A novel polypeptide of the invention can also possess one or more of the following functions or properties: (1) an activator functioning to regulate one or more genes by increasing the rate of transcription, (2) an activator functioning to positively modulate an allosteric enzyme, (3) an adaptor functioning to sort cargo molecules into transport vesicles, (4) an

adaptor functioning to form a clathrin-coated vesicle, (5) an adhesion molecule functioning to mediate the adhesion of cells with other cells and/or the extracellular matrix, (6) an ATPase functioning to move ions or small molecules across a membrane against a chemical concentration gradient or electrical potential, (7) an ATPase functioning to translocate nucleotides across membranes, (8) a breakpoint-related sequence functioning as an oncoprotein, (9) a breakpoint-related sequence functioning as a tumor-specific antigen, (10) a channel functioning as a water channel, (11) a channel functioning as an ion channel, (12) a checkpoint-related sequence functioning at DNA damage checkpoints, (13) a checkpoint-related sequence functioning at replication checkpoints, (14) a checkpoint-related sequence functioning to initiate signal transduction cascades eliciting cell cycle arrest, DNA repair, or apoptosis, (15) a complex functioning as a protein scaffold, (16) a complex functioning in ADP-ribosylation, (17) a dehydrogenase functioning to synthesize amino acids, (18) a disintegrin functioning to inhibit blood clotting, (19) a disintegrin functioning as a metallopeptidase, (20) a GTPase functioning as a negative regulator of p53, (21) a GTPase functioning to stimulate ras GTPase activity, (22) a helicase functioning in DNA replication, (23) a hydrolase functioning in proprionate metabolism, (24) an integrase functioning to integrate a DNA copy of a retroviral genome into a host chromosome, (25) an integrin functioning as a tumor marker, (26) an integrin functioning in cell migration, (27) an isomerase functioning as an immunosuppressant, (28) a membrane protein functioning as a scaffolding component at the cytoplasmic face of a lipid raft, (29) a membrane protein functioning as a ligand for a receptor tyrosine kinase, (30) oxygenases and peroxidases functioning as antioxidants, (31) a phospholipase functioning in eicosanoid synthesis, (32) a phospholipase functioning in preserving the intestinal mucosa, (33) a prosaposin functioning in lipid catabolism, (34) a proteasome component functioning in muscle wasting, (35) a reductase-related sequence functioning as a coenzyme A reductase inhibitor, (36) a reverse transcriptase functioning as an RNA-dependent reverse transcriptase, (37) a reverse transcriptase functioning as a DNA-dependent reverse transcriptase, (38) an RNase functioning in viral assembly, (39) an RNase H functioning to form oligonucleotides that prime DNA synthesis, (40) an RNase H functioning to cleave the RNA strand of an RNA-DNA hybrid, (41) SH3 domains functioning in actin cytoskeletal organization, (42) SH3 domains functioning in signal transduction, (43) a synthetase functioning as an autoantigen (44) synthetases

functioning in nucleotide sugar phosphate synthesis, (45) TATA boxes functioning as a transcription initiators, (46) tat functioning as a transcriptional coactivator, (47) transferases functioning in signal transduction, (48) transposases functioning as gene transfer agents, (49) ubiquitins functioning to protect cells against tumor necrosis factor induced cell death, (50) proteasome components and ubiquitin functioning in protein degradation, (51) a virus-related sequence functioning to confer resistance to infection by viruses, (52) other sequences of the invention interacting with one or more proteins, (53) other sequences of the invention enzymatically modifying one or more proteins, (54) other sequences of the invention binding one or more small molecule ligands, (55) other sequences of the invention binding one or more peptides, (56) other sequences of the invention binding one or more carbohydrates, and (57) other sequences of the invention functioning in vesicular transport.

[0398] In some embodiments, the present novel polypeptide modulates the cells or tissues of animals, particularly humans, such as, for example, by stimulating, enhancing or inhibiting T or B cell function or the function of other hematopoeitic cells or bone marrow cells; modulates adult or embryonic stem cell or precursor cell growth or differentiation; modulates cell function or activity of neuronal cells or other cells of the CNS, heart cells, liver cells, kidney cells, lung cells, pancreatic cells, gastrointestinal cells, spleen cells, breast cells, prostate cells, ovarian cells, and the like.

[0399] In some embodiments, a subject polypeptide is present as a multimer. Multimers include homodimers, homotrimers, homotetramers, and multimers that include more than four monomeric units. Multimers also include heteromultimers, e.g., heterodimers, heterotrimers, heterotetramers, etc. where the subject polypeptide is present in a complex with proteins other than the subject polypeptide. Where the multimer is a heteromultimer, the subject polypeptide can be present in a 1:1 ratio, a 1:2 ratio, a 2:1 ratio, or other ratio, with the other protein(s).

[0400] In addition to the above specifically listed proteins, polypeptides from other species are also provided, including mammals, such as: primates, rodents, e.g., mice, rats, hamsters, guinea pigs; domestic animals, e.g., sheep, pig, horse, cow, goat, rabbit, dog, cat; and humans, as well as non-mammalian species, e.g., avian, reptile and amphibian, insect, crustacean, fish, plant, fungus, and protozoa.

[0401] By "homolog" is meant a protein having at least about 35 %, at least about 40%, at least about 60%, at least about 70%, at least about 75%, at least about

80%, at least about 85%, at least about 90%, or at least about 95%, or higher, amino acid sequence identity to the reference polypeptide, as measured with the "GAP" program (part of the Wisconsin Sequence Analysis Package available through the Genetics Computer Group, Inc. (Madison WI)), where the parameters are: Gap weight:12; length weight:4. In many embodiments of interest, homology will be at least about 75%, at least about 80%, or at least 85%, where in certain embodiments of interest, homology will be as high as about 90%.

[0402] Also provided are polypeptides that are substantially identical to the at least one amino acid sequence shown in the Sequence Listing, or a fragment thereof, whereby substantially identical is meant that the protein has an amino acid sequence identity to the reference sequence of at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 97%, at least about 98%, or at least about 99%.

[0403] The proteins of the subject invention (e.g., polypeptides encoded by the nucleotide sequences shown in SEQ ID NOS.: 1 - 209 and 419 - 627, and polypeptide sequences shown in SEQ ID NOS.: 210 - 418) have been separated from their naturally occurring environment and are present in a non-naturally occurring environment. In certain embodiments, the proteins are present in a composition where they are more concentrated than in their naturally occurring environment. For example, purified polypeptides are provided.

[0404] In addition to naturally occurring proteins, polypeptides that vary from naturally occurring forms are also provided. Fusion proteins can comprise a subject polypeptide, or fragment thereof, and a polypeptide other than a subject polypeptide ("the fusion partner") fused in-frame at the N-terminus and/or C-terminus of the subject polypeptide, or internally to the subject polypeptide.

[0405] Suitable fusion partners include, but are not limited to, immunologically detectable proteins (e.g., epitope tags, such as hemagglutinin, FLAG, and c-myc); polypeptides that provide a detectable signal or that serve as detectable markers (e.g., a fluorescent protein, e.g., a green fluorescent protein, a fluorescent protein from an Anthozoan species; β-galactosidase; luciferase; cre recombinase; and the like); polypeptides that provide a catalytic function or induce a cellular response; polypeptides that provide for secretion of the fusion protein from a eukaryotic cell; polypeptides that provide for secretion of the fusion protein from a prokaryotic cell; polypeptides that provide for binding to metal ions (e.g., His_n, where

n = 3-10, e.g., 6His) and structural proteins. Fusion partners can also be those that are able to stabilize the present polypeptide, such as polyethylene glycol ("PEG") and a fragment of an immunoglobulin, such as the Fc fragment of IgG, IgE, IgA, IgM, and/or IgD.

[0406] Detection methods are chosen based on the detectable fusion partner. For example, where the fusion partner provides an immunologically recognizable epitope, an epitope-specific antibody can be used to quantitatively detect the level of polypeptide. In some embodiments, the fusion partner provides a detectable signal, and in these embodiments, the detection method is chosen based on the type of signal generated by the fusion partner. For example, where the fusion partner is a fluorescent protein, fluorescence is measured.

[0407] Where the fusion partner is an enzyme that yields a detectable product, the product can be detected using an appropriate means. For example, β-galactosidase can, depending on the substrate, yield a colored product that can be detected with a spectrophotometer, and the fluorescent protein luciferase can yield a luminescent product detectable with a luminometer.

[0408] In some embodiments, a polypeptide of the invention comprises at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, at least about 300, at least about 350, at least about 400, at least about 450, at least about 500, at least about 550, at least about 600, at least about 650, at least about 700, at least about 750, at least about 800, at least about 850, at least about 900, at least about 950, or at least about 1000 contiguous amino acid residues of at least one of the sequences according to SEQ ID NOS.: 210 - 418, up to and including the entire amino acid sequence.

[0409] Fragments of the subject polypeptides, as well as polypeptides comprising such fragments, are also provided. Fragments of polypeptides of interest will typically be at least about 5, at least about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, at least about 50, at least about 75, at least about 100, at least about 150, at least about 200, at least about 250, or at least 300 aa in length or longer, where the fragment will have a stretch of amino acids that is identical to the subject protein of at least about 5, at least

about 8, at least about 10, at least about 15, at least about 18, at least about 20, at least about 25, at least about 30, or at least about 50 aa in length.

[0410] In some embodiments, fragments exhibit one or more activities associated with a corresponding naturally occurring polypeptide. Fragments find utility in generating antibodies to the full-length polypeptide; and in methods of screening for candidate agents that bind to and/or modulate polypeptide activity. Specific fragments of interest include those with enzymatic activity, those with biological activity including the ability to serve as an epitope or immunogen, and fragments that bind to other proteins or to nucleic acids.

[0411] The invention provides polypeptides comprising such fragments, including, e.g., fusion polypeptides comprising a subject polypeptide fragment fused in frame (directly or indirectly) to another protein (the "fusion partner"), such as the signal peptide of one protein being fused to the mature polypeptide of another protein. Such fusion proteins are typically made by linking the encoding polynucleotides together in a vector or cassette. Suitable fusion partners include, but are not limited to, immunologically detectable proteins (e.g., epitope tags, such as hemagglutinin, FLAG, and c-myc); polypeptides that provide a detectable signal or that serve as detectable markers (e.g., a fluorescent protein, e.g., a green fluorescent protein, a fluorescent protein from an Anthozoan species; β-galactosidase; luciferase; cre recombinase); polypeptides that provide a catalytic function or induce a cellular response; polypeptides that provide for secretion of the fusion protein from a eukaryotic cell; polypeptides that provide for secretion of the fusion protein from a prokaryotic cell; polypeptides that provide for binding to metal ions (e.g., His_n, where n = 3-10, e.g., 6His) and structural proteins. Fusion partners can also be those that are able to stabilize the present polypeptide, such as polyethylene glycol ("PEG") and a fragment of an immunoglobulin, such as the Fc fragment of IgG, IgE, IgA, IgM, and/or IgD.

Polypeptide Preparation.

[0412] Polypeptides of the invention can be obtained from naturally-occurring sources or produced synthetically. The sources of naturally occurring polypeptides will generally depend on the species from which the protein is to be derived, i.e., the proteins will be derived from biological sources that express the proteins. The subject proteins can also be derived from synthetic means, e.g., by expressing a recombinant gene encoding a protein of interest in a suitable system or

host or enhancing endogenous expression, as described in more detail above. Further, small peptides can be synthesized in the laboratory by techniques well known in the art.

[0413] In all cases, the product can be recovered by any appropriate means known in the art. For example, convenient protein purification procedures can be employed (e.g., see Guide to Protein Purification, Deuthscher et al., 1990). That is, a lysate can be prepared from the original source, (e.g., a cell expressing endogenous polypeptide, or a cell comprising the expression vector expressing the polypeptide(s)), and purified using HPLC, exclusion chromatography, gel electrophoresis, or affinity chromatography, and the like.

[0414] The invention thus also provides methods of producing polypeptides. Briefly, the methods generally involve introducing a nucleic acid construct into a host cell *in vitro* and culturing the host cell under conditions suitable for expression, then harvesting the polypeptide, either from the culture medium or from the host cell, (e.g., by disrupting the host cell), or both, as described in detail above. The invention also provides methods of producing a polypeptide using cell-free *in vitro* transcription/translation methods, which are well known in the art, also as provided above

[0415] Moreover, the invention provides polypeptides, including polypeptide fragments, as targets for therapeutic intervention, including use in screening assays, for identifying agents that modulate polypeptide level and/or activity, and as targets for antibody and small molecule therapeutics, for example, in the treatment of disorders.

Kits

[0416] The present invention provides kits for diagnosing disease states based on the detected presence and/or level of polynucleotide or polypeptide in a biological sample, and/or the detected presence and/or level of biological activity of the polynucleotide or polypeptide. The invention further provides kits for detecting the presence and/or a level of a polynucleotide or polypeptide in a biological sample and/or or the detected presence and/or level of biological activity of the polynucleotide or polypeptide. Procedures using these kits can be performed by clinical laboratories, experimental laboratories, medical practitioners, or private individuals.

[0417] The kits of the invention will comprise a molecule of the invention. The kits for detecting a polynucleotide will also comprise a moiety that specifically hybridizes to a polynucleotide of the invention. The polynucleotide molecule can be of any length. For example, it can comprise a polynucleotide of at least 6, at least 7, at least 8, or at least 9 contiguous nucleotides of a molecule of the invention. Kits of the invention for detecting a subject polypeptide will comprise a moiety that specifically binds to a polypeptide of the invention; the moiety includes, but is not limited to, a polypeptide-specific antibody.

[0418] The kits are useful in diagnostic applications. For example, the kit is useful to determine whether a given DNA sample isolated from an individual comprises an expressed nucleic acid, a polymorphism, or other variant.

[0419] Kits for detecting polynucleotides comprise a pair of nucleic acids in a suitable storage medium, e.g., a buffered solution, in a suitable container. The pair of isolated nucleic acid molecules serve as primers in an amplification reaction (e.g., a polymerase chain reaction). The kit can further include additional buffers, reagents for polymerase chain reaction (e.g., deoxynucleotide triphosphates (dNTP), a thermostable DNA polymerase, a solution containing Mg^{2+} ions (e.g., $MgCl_2$), and other components well known to those skilled in the art for carrying out a polymerase chain reaction). The kit can further include instructions for use, which may be provided in a variety of forms, e.g., printed information, or compact disc, and the like. The kit may further include reagents necessary to extract DNA from a biological sample and reagents for generating a cDNA copy of an mRNA. The kit may optionally provide additional useful components, including, but not limited to, buffers, developing reagents, labels, reacting surfaces, means for detections, control samples, standards, and interpretive information.

[0420] In some embodiments, a kit of the invention for detecting a polynucleotide, such as an mRNA encoding a polypeptide, comprises a pair of nucleic acids that function as "forward" and "reverse" primers that specifically amplify a cDNA copy of the mRNA. The "forward" and "reverse" primers are provided as a pair of isolated nucleic acid molecules, each from about 10 to about 200 nucleotides in length, the first nucleic acid molecule of the pair comprising a sequence of at least about 10 contiguous nucleotides having 100% sequence identity to a nucleic acid sequence shown in from SEQ ID NOS.: 1 - 209 and 419 - 627, and the second nucleic acid molecule of the pair comprising a sequence of at least about 10 contiguous

nucleotides having 100% sequence identity to the reverse complement of a nucleic acid sequence shown in SEQ ID NOS.: 1 - 209 and 419 - 627 , wherein the sequence of the second nucleic acid molecule is located 3' of the nucleic acid sequence of the first nucleic acid molecule. The primer nucleic acids are prepared using any known method, e.g., automated synthesis. In some embodiments, one or both members of the pair of nucleic acid molecules comprise a detectable label. The kit may include blocking reagents, buffers, and reagents for developing and/or detecting the detectable label. The kit may also include instructions for use, controls, and interpretive information.

[0421] Where the kit provides for detecting enzymatic activity, it includes a substrate that provides for a detectable product when acted upon by a polypeptide of interest. The kit may further include reagents necessary to detect and develop the detectable marker.

[0422] The present invention provides for kits with unit doses of an active agent. These agents are described in more detail below. In some embodiments, the agent is provided in oral or injectable doses. Such kits will comprise containers containing the unit doses and an informational package insert describing the use and attendant benefits of the drugs in treating a condition of interest.

Tables

Table 1. Characteristics of the Claimed Sequences, and of the Protein With the Highest Degree of Similarity to Each

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000569N0_160000_gene_prediction1	100	0.24	gi|28829824|gb|AAO52326.1|	similar to Dictyostelium discoideum (Slime mold). Protein kinase
HG1001052N0_0_gene_prediction1	41	0.46	gi|28918732|gb|EAA28402.1|	predicted protein [Neurospora crassa] gi|29150136|emb|CAD79696.1| hypothetical protein [Neurospora
HG1000498N0_160000_gene_prediction1	130	0.16	gi|28921639|gb|EAA30925.1|	predicted protein [Neurospora crassa]
HG1000685N0_160000_gene_prediction1	370	0.12	gi|22770596|gb|AAN06673.1|	voltage-gated calcium channel alpha(2)delta-3 subunit [Homo sapiens]
HG1000622N0_160000_gene_prediction1	423	0.49	gi|106322|pir||B34087	hypothetical protein (L1H 3' region) - human
HG1000390N0_1000_gene_prediction1	515	1	gi|22044951|ref|XP_087331.5|	similar to URB [Homo sapiens]
HG1000806N0_20000_gene_prediction1	92	0.82	gi|23592855|ref|XP_129487.2|	hypothetical protein MGC40674 [Mus musculus]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1001489N0_20000_gene_prediction1	92	0.82	gi|23592855|ref|XP_129487.2|	hypothetical protein MGC40674 [Mus musculus]
HG1001478N0_10000_gene_prediction1	210	0.13	gi|23475363|ref|ZP_00130651.1|	hypothetical protein [Desulfovibrio desulfuricans G20]
HG1000806N0_160000_gene_prediction1	146	0.52	gi|23592855|ref|XP_129487.2|	hypothetical protein MGC40674 [Mus musculus]
HG1000403N0_160000_gene_prediction1	83	0.3	gi|16124876|ref|NP_419440.1|	conserved hypothetical protein [Caulobacter crescentus CB15]
HG1001201N0_160000_gene_prediction1	276	0.12	gi|28506221|ref|XP_282957.1|	RIKEN cDNA D130005J21 gene [Mus musculus]
HG1000617N0_20000_gene_prediction1	78	0.25	gi|23501223|ref|NP_697350.1|	sensor histidine kinase/response regulator [Brucella suis 1330]
HG1001334N0_160000_gene_prediction1	243	0.16	gi|14329731|emb|CAC40671.1|	high molecular weight glutenin subunit x [Secale cereale]
HG1000834N0_160000_gene_prediction1	81	0.39	gi|2072955|gb|AAC51265.1|	p40 [Homo sapiens]
HG1000752N0_10000_gene_prediction1	119	0.15	gi|16761576|ref|NP_457193.1|	putative transcriptional regulator [Salmonella enterica subsp. enterica serovar Typhi]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000839N0_160000_gene_prediction2	412	1	gi|19923957|ref|NP_612440.1|	hypothetical protein BC011982 [Homo sapiens]
HG1000360N0_20000_gene_prediction1	494	0.14	gi|6322209|ref|NP_012284.1|	Required for invasion and pseudohyphae formation in response to nitrogen starvation; Muc1p
HG1000360N0_20000_gene_prediction2	469	0.15	gi|6322209|ref|NP_012284.1|	Required for invasion and pseudohyphae formation in response to nitrogen starvation; Muc1p
HG1000559N0_10000_gene_prediction1	46	0.43	gi|18557931|ref|XP_087525.1|	similar to PRO1546 [Homo sapiens]
HG1000570N0_160000_gene_prediction1	146	0.17	gi|17543234|ref|NP_501173.1|	Putative protein, nematode specific [Caenorhabditis elegans]
HG1000617N0_40000_gene_prediction1	413	0.52	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1000615N0_160000_gene_prediction2	77	1	gi|22061362|ref|XP_171933.1|	similar to ribosomal protein S4, cytosolic [validated] - rat [Homo sapiens]
HG1000617N0_160000_gene_prediction1	391	0.55	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1000621N0_160000_gene_prediction2	77	1	gi|22061362|ref|XP_171933.1|	similar to ribosomal protein S4, cytosolic [validated] - rat [Homo sapiens]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1001502N0_160000_gene_prediction2	257	0.9	gi|2136258|pir||I59377	template activating factor-I, splice form alpha - human
HG1000501N0_160000_gene_prediction1	143	0.23	gi|24664732|ref|NP_648789.2|	CG7372-PA [Drosophila melanogaster] gi|7294254|gb|AAF49605.1| CG7372-PA [Drosophila melanogaster]
HG1000656N0_10000_gene_prediction1	81	0.24	gi|15608619|ref|NP_215997.1|	hypothetical protein Rv1481 [Mycobacterium tuberculosis H37Rv]
HG1000656N0_10000_gene_prediction2	81	0.24	gi|15608619|ref|NP_215997.1|	hypothetical protein Rv1481 [Mycobacterium tuberculosis H37Rv]
HG1000750N0_160000_gene_prediction1	184	0.11	gi|27469362|ref|NP_010782.2|	Ribosomal Small subunit Mitochondria; Rsm28p [Saccharomyces cerevisiae]
HG1001012N0_160000_gene_prediction1	219	0.74	gi|22325370|ref|NP_680544.1|	MGC29635 protein [Homo sapiens] gi|22325372|ref|NP_680545.1| MGC29635 protein [Homo sapiens]
HG1001237N0_10000_gene_prediction1	50			no_blastp_hit
HG1000390N0_160000_gene_prediction1	1885	0.49	gi|22044951|ref|XP_087331.5|	similar to URB [Homo sapiens]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000847N0_10000_gene_prediction1	72			no_blastp_hit
HG1000331N0_160000_gene_prediction1	281	1	gi|27499231|ref|XP_166856.3|	similar to SLIT1-Sa splicing product [Rattus norvegicus] [Homo sapiens]
HG1000391N0_160000_gene_prediction2	199	0.16	gi|19553189|ref|NP_601191.1|	translation initiation factor 2 [Corynebacterium glutamicum ATCC 13032]
HG1000597N0_160000_gene_prediction1	77	0.75	gi|25518567|pir||E86336	hypothetical protein F14O10.11 - Arabidopsis thaliana
HG1000415N0_10000_gene_prediction1	149	0.18	gi|13959539|sp|O15198|SMA9_HUMAN	Mothers against decapentaplegic homolog 9 (SMAD 9) (Mothers against DPP homolog 9)
HG1000618N0_10000_gene_prediction1	96	0.26	gi|25020268|ref|XP_207842.1|	hypothetical protein XP_207842 [Mus musculus]
HG1001197N0_160000_gene_prediction1	200	0.09	gi|27498170|ref|XP_209693.1|	similar to hypothetical protein [Macaca fascicularis] [Homo sapiens]
HG1000599N0_5000_gene_prediction1	124	0.57	gi|23509612|ref|NP_702279.1|	hypothetical protein [Plasmodium falciparum 3D7]
HG1000424N0_5000_gene_prediction1	167	0.32	gi|25020846|ref|XP_207573.1|	similar to hypothetical protein [Plasmodium yoelii yoelii] [Mus musculus]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1001485N0_5000_gene_prediction1	144	0.66	gi|26327365|dbj|BAC27426.1|	unnamed protein product [Mus musculus]
HG1000674N0_160000_gene_prediction1	25	0.56	gi|5922609|dbj|BAA84610.1|	unnamed protein product [Oryza sativa (japonica cultivar-group)]
HG1000339N0_160000_gene_prediction1	169	0.09	gi|18603697|ref|XP_085597.1|	hypothetical protein XP_085597 [Homo sapiens]
HG1000340N0_160000_gene_prediction1	421	0.99	gi|19923485|ref|NP_057508.2|	cisplatin resistance-associated overexpressed protein [Homo sapiens]
HG1000344N0_160000_gene_prediction1	71	0.23	gi|15893930|ref|NP_347279.1|	Conserved domain seen in the bacterial SpoT [Clostridium acetobutylicum]
HG1000365N0_20000_gene_prediction1	92	0.78	gi|4504255|ref|NP_002097.1|	H2A histone family, member Z; H2AZ histone [Homo sapiens]
HG1000384N0_160000_gene_prediction1	469	0.54	gi|6624128|gb|AAF19255.1|AC004858_3	U1 small ribonucleoprotein 1SNRP homolog; match to PID:g4050087 [Homo sapiens]
HG1000448N0_160000_gene_prediction1	147	0.25	gi|18087335|gb|AAL58838.1|AF390028_1	serine/threonine protein kinase kkialre-like 1 [Homo sapiens]
HG1000506N0_160000_gene_prediction1	204	0.43	gi|25031822|ref|XP_207741.1|	hypothetical protein XP_207741 [Mus musculus]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000550N0_160000_gene_prediction1	469	0.54	gi|6624128|gb|AAF19255.1|AC004858_3	U1 small ribonucleoprotein 1SNRP homolog; match to PID:g4050087 [Homo sapiens]
HG1000647N0_160000_gene_prediction1	145	0.15	gi|23056247|gb|ZP_00082297.1|	hypothetical protein [Geobacter metallireducens]
HG1000688N0_160000_gene_prediction1	176	0.76	gi|21903371|sp|O15072|ATS3_HUMAN	ADAMTS-3 precursor (A disintegrin and metalloproteinase with thrombospondin motifs 3)
HG1000904N0_160000_gene_prediction1	332	0.82	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1001417N0_5000_gene_prediction1	1241	0.13	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4 regulatory subunit 2 [Homo sapiens]
HG1001485N0_160000_gene_prediction1	239	0.58	gi|10047201|dbj|BAB13394.1|	KIAA1568 protein [Homo sapiens]
HG1001502N0_160000_gene_prediction1	1056	0.95	gi|2209278|gb|AAB66673.1|	oxytocinase splice variant 2 [Homo sapiens]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000343N0_160000_gene_prediction1	440	0.53	gi|21750183|dbj|BAC03736.1|	unnamed protein product [Homo sapiens]
HG1000343N0_160000_gene_prediction2	56	0.23	gi|24585773|ref|NP_724384.1|	CG6448-PA [Drosophila melanogaster] gi|24585775|ref|NP_610134.2| CG6448-PB [Drosophila melanogaster]
HG1000369N0_160000_gene_prediction1	478	0.04	gi|2120169|pir||S12598	gene X protein - hepatitis B virus (subtype adr)
HG1000378N0_160000_gene_prediction1	243	0.19	gi|5802824|gb|AAD51799.1|AF164615_1	Gag-Pro-Pol protein [Homo sapiens]
HG1000387N0_160000_gene_prediction1	188	0.35	gi|225047|prf||1207289A	reverse transcriptase related protein
HG1000387N0_160000_gene_prediction2	133	0.27	gi|18557931|ref|XP_087525.1|	similar to PRO1546 [Homo sapiens]
HG1000408N0_160000_gene_prediction1	69	0.23	gi|12229657|sp|Q9LU41|ACA9_ARATH	Potential calcium-transporting ATPase 9, plasma membrane-type (Ca2+-ATPase, isoform 9)
HG1000431N0_160000_gene_prediction1	180	0.28	gi|27482385|ref|XP_208702.1|	hypothetical protein XP_208702 [Homo sapiens]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000457N0_160000_gene_prediction1	118	0.4	gi|28485359|ref|XP_197468.2|	hypothetical protein XP_197468 [Mus musculus]
HG1000458N0_160000_gene_prediction1	256	0.2	gi|27704378|ref|XP_230791.1|	similar to mannosidase, beta A, lysosomal-like [Homo sapiens] [Rattus norvegicus]
HG1000463N0_160000_gene_prediction1	156	0.23	gi|27721443|ref|XP_213140.1|	similar to cofilin - rat [Rattus norvegicus]
HG1000463N0_160000_gene_prediction2	113	0.17	gi|21289376|gb|EAA01669.1|	agCP11939 [Anopheles gambiae str. PEST]
HG1000481N0_160000_gene_prediction1	130	1	gi|17436547|ref|XP_067994.1|	similar to heat shock factor binding protein 1 [Homo sapiens]
HG1000592N0_160000_gene_prediction1	1036	0.38	gi|14017873|dbj|BAB47457.1|	KIAA1828 protein [Homo sapiens]
HG1000608N0_160000_gene_prediction1	128	0.22	gi|14249206|ref|NP_116044.1|	hypothetical protein MGC10997 [Homo sapiens]
HG1000615N0_160000_gene_prediction1	132	0.18	gi|22977434|gb|ZP_00023237.1|	hypothetical protein [Ralstonia metallidurans]
HG1000621N0_160000_gene_prediction1	188	0.35	gi|225047|prf||1207289A	reverse transcriptase related protein

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000621N0_160000_gene_prediction3	133	0.27	gi|18557931|ref|XP_087525.1|	similar to PRO1546 [Homo sapiens]
HG1000652N0_160000_gene_prediction1	110	0.62	gi|337663|gb|AAA36590.1|	ORF1; putative
HG1000663N0_160000_gene_prediction1	335	0.06	gi|9392624|gb|AAF87223.1|AF247180_1	sialic acid-binding immunoglobulin-like lectin-9 [Homo sapiens]
HG1000700N0_160000_gene_prediction1	120	0.4	gi|9967319|dbj|BAB12359.1|	hypothetical protein [Macaca fascicularis]
HG1000701N0_160000_gene_prediction1	322	0.09	gi|10140726|gb|AAG13559.1|AC073867_5	putative proteophosphoglycan [Oryza sativa]
HG1000709N0_160000_gene_prediction1	168	0.2	gi|138692|sp|P22856|VL96_IRV1	L96 protein gi|93562|pir||JH0225 L96 protein - Tipula iridescent virus
HG1000720N0_160000_gene_prediction1	179	0.21	gi|9910320|ref|NP_064473.1|	neurestin alpha [Rattus norvegicus] gi|5712201|gb|AAD47383.1|AF086607_1 neurestin alpha [Rattus
HG1000786N0_160000_gene_prediction2	85	0.24	gi|17556188|ref|NP_497536.1|	Elongation factor Tu domain 2 containing protein (133.3 kD) [Caenorhabditis elegans]
HG1000848N0_160000_gene_prediction1	242	0.42	gi|25527052|pir||JC7766	three-amino-acid loop extension(TALE) homeodomain protein, PKNOX2 - human

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000904N0_160000_gene_prediction3	70	0.24	gi|21398270|ref|NP_654255.1|	aminotran_3, Aminotransferase class-III [Bacillus anthracis A2012]
HG1001214N0_20000_gene_prediction1	54			no_blastp_hit
HG1001229N0_160000_gene_prediction1	355	0.94	gi|106322|pir||B34087	hypothetical protein (L1H 3' region) - human
HG1001468N0_160000_gene_prediction1	288	0.11	gi|16041152|dbj|BAB69743.1|	hypothetical protein [Macaca fascicularis]
HG1000382N0_160000_gene_prediction1	1609	0.85	gi|4505017|ref|NP_002327.1|	low density lipoprotein receptor-related protein 6; low density lipoprotein-related protein 6
HG1000591N0_160000_gene_prediction1	1895	0.85	gi|22779441|dbj|BAC15608.1|	FELE-2 [Homo sapiens]
HG1000904N0_160000_gene_prediction4	1084	0.93	gi|87211|pir||PL0009	complement C3d/Epstein-Barr virus receptor precursor - human
HG1000267N0_5000_gene_prediction1	153	0.9	gi|7669492|ref|NP_002037.2|	glyceraldehyde-3-phosphate dehydrogenase [Homo sapiens]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000332N0_10000_gene_prediction1	168	0.11	gi|16550911|dbj|BAB71077.1|	unnamed protein product [Homo sapiens]
HG1000341N0_5000_gene_prediction1	167	0.12	gi|27674743|ref|XP_239301.1|	hypothetical protein XP_239301 [Rattus norvegicus]
HG1000341N0_10000_gene_prediction1	195	0.17	gi|7715984|gb|AAF68235.1|AF206244_1	seroreactive antigen BMN1-2 [Babesia microti]
HG1000353N0_160000_gene_prediction1	120	0.22	gi|28829049|gb|AAO51624.1|	similar to Dictyostelium discoideum (Slime mold). LvsB
HG1000359N0_160000_gene_prediction1	560	0.46	gi|25046852|ref|XP_207505.1|	similar to endonuclease/reverse transcriptase [Mus musculus]
HG1000363N0_160000_gene_prediction1	138	0.22	gi|29135033|ref|NP_803663.1|	PHIKZ097 [Pseudomonas phage phiKZ] gi|18996562|gb|AAL82998.1|AF399011_97 PHIKZ097 [Pseudomonas phage
HG1000364N0_160000_gene_prediction1	436	0.17	gi|7510074|pir||T31611	hypothetical protein Y50E8A.g - Caenorhabditis elegans
HG1000367N0_160000_gene_prediction1	114	0.12	gi|1363528|pir||S58484	gag protein - maize gi|507845|gb|AAA93147.1| gag gene product
HG1000379N0_160000_gene_prediction1	336	0.83	gi|2072951|gb|AAC51263.1|	putative p150 [Homo sapiens]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000390N0_10000_gene_prediction1	928	0.65	gi|22044951|ref|XP_087331.5|	similar to URB [Homo sapiens]
HG1000390N0_5000_gene_prediction1	676	0.79	gi|22044951|ref|XP_087331.5|	similar to URB [Homo sapiens]
HG1000391N0_160000_gene_prediction1	254	0.25	gi|24061772|gb|AAN39840.1|	early B-cell factor-associated zinc finger protein; 30 kruppel-like zinc finger protein [Mus
HG1000407N0_160000_gene_prediction1	389	0.98	gi|6176586|sp|P51026|YI2A_ECOLI	Insertion element IS2A hypothetical 48.2 kDa protein
HG1000408N0_160000_gene_prediction2	180	0.15	gi|27682897|ref|XP_214580.1|	protocadherin alpha 3 [Rattus norvegicus]
HG1000429N0_160000_gene_prediction1	189	0.11	gi|20883959|ref|XP_140527.1|	similar to neuropilin and tolloid like-1 [Mus musculus]
HG1000431N0_20000_gene_prediction1	250	0.2	gi|10436424|dbj|BAB14833.1|	unnamed protein product [Homo sapiens]
HG1000435N0_160000_gene_prediction1	251	0.16	gi|25031481|ref|XP_207546.1|	similar to hypothetical protein [Plasmodium yoelii yoelii] [Mus musculus]
HG1000474N0_5000_gene_prediction1	82	0.26	gi|14755300|ref|XP_042833.1|	similar to KIAA0295 [Homo sapiens]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000489N0_160000_gene_prediction1	497	0.88	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1000499N0_160000_gene_prediction1	189	0.13	gi|27702816|ref|XP_224411.1|	similar to hypothetical protein [Plasmodium yoelii yoelii] [Rattus norvegicus]
HG1000500N0_160000_gene_prediction1	564	0.4	gi|27482474|ref|XP_068681.3|	similar to seven transmembrane helix receptor [Homo sapiens]
HG1000505N0_160000_gene_prediction1	322	0.07	gi|23502549|ref|NP_698676.1|	cell division protein FtsH [Brucella suis 1330]
HG1000509N0_10000_gene_prediction1	82	0.26	gi|14755300|ref|XP_042833.1|	similar to KIAA0295 [Homo sapiens]
HG1000519N0_160000_gene_prediction1	344	0.11	gi|21264565|ref|NP_006006.3|	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin f1 isoform a;
HG1000521N0_160000_gene_prediction1	249	0.11	gi|25032631|ref|XP_194529.1|	similar to cDNA sequence AY036118 [Mus musculus]
HG1000549N0_160000_gene_prediction1	299	0.95	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1000549N0_160000_gene_prediction2	207	0.12	gi|23055534|gb|ZP_00081634.1|	hypothetical protein [Geobacter metallireducens]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000549N0_160000_gene_prediction3	299	0.95	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1000562N0_160000_gene_prediction1	362	0.16	gi|28494166|ref|XP_286229.1|	expressed sequence AU043053 [Mus musculus]
HG1000582N0_160000_gene_prediction1	1021	0.7	gi|24307949|ref|NP_055099.1|	neurochondrin [Homo sapiens] gi|4887649|dbj|BAA77830.1| neurochondrin-1 [Homo sapiens]
HG1000598N0_160000_gene_prediction1	682	1	gi|4885419|ref|NP_005332.1|	GLI-Kruppel family member HKR3 [Homo sapiens]
HG1000606N0_20000_gene_prediction1	185	0.1	gi|3253147|gb|AAC24492.1|	CMADS1 [Ceratopteris richardii]
HG1000607N0_160000_gene_prediction1	417	0.48	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1000608N0_20000_gene_prediction1	96	0.3	gi|14249206|ref|NP_116044.1|	hypothetical protein MGC10997 [Homo sapiens]
HG1000616N0_1000_gene_prediction1	20			no_blastp_hit
HG1000622N0_160000_gene_prediction2	79	0.17	gi|28898213|ref|NP_797818.1|	DnaK-related protein [Vibrio parahaemolyticus RIMD 2210633]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000625N0_160000_gene_prediction1	642	0.49	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1000628N0_40000_gene_prediction1	172	0.16	gi|25029823|ref|XP_207224.1|	similar to Transposase of Tn10 [Oryza sativa] [Mus musculus]
HG1000628N0_20000_gene_prediction1	577	0.14	gi|2072972|gb|AAC51276.1|	putative p150 [Homo sapiens]
HG1000638N0_5000_gene_prediction1	38			no_blastp_hit
HG1000642N0_160000_gene_prediction1	929	0.95	gi|5070622|gb|AAD39215.1|AF148856_2	unknown [Homo sapiens]
HG1000646N0_160000_gene_prediction1	113	0.2	gi|22962666|gb|ZP_00010272.1|	hypothetical protein [Rhodopseudomonas palustris]
HG1000649N0_160000_gene_prediction1	275	0.9	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1000652N0_160000_gene_prediction2	145	1	gi|27498807|ref|XP_208381.1|	similar to nuclear receptor binding factor-2 [Homo sapiens]
HG1000656N0_160000_gene_prediction1	244	0.08	gi|15608619|ref|NP_215997.1|	hypothetical protein Rv1481 [Mycobacterium tuberculosis H37Rv]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000656N0_160000_gene_prediction2	244	0.08	gi|15608619|ref|NP_215997.1|	hypothetical protein Rv1481 [Mycobacterium tuberculosis H37Rv]
HG1000670N0_160000_gene_prediction1	1306	0.89	gi|4759030|ref|NP_004251.1|	RecQ protein-like 4; RecQ protein 4 [Homo sapiens]
HG1000685N0_160000_gene_prediction2	58	0.31	gi|15242074|ref|NP_200535.1|	salt-inducible protein-like; protein id: At5g57250.1 [Arabidopsis thaliana]
HG1000700N0_160000_gene_prediction2	231	1	gi|22538495|ref|NP_071924.1|	Williams Beuren syndrome chromosome region 17; polypeptide N-acetylgalactosaminyltransferase
HG1000738N0_160000_gene_prediction1	157	0.23	gi|6912492|ref|NP_036456.1|	mitogen-activated protein kinase 8 interacting protein 2 isoform 1; PRKM8 interacting
HG1000739N0_160000_gene_prediction1	218	0.97	gi|5070621|gb|AAD39214.1|AF148856_1	unknown [Homo sapiens]
HG1000739N0_160000_gene_prediction2	721	0.45	gi|17980446|gb|AAL50636.1|	unknown [Homo sapiens]
HG1000779N0_160000_gene_prediction1	80	0.31	gi|23308823|ref|NP_600311.2|	hypothetical protein [Corynebacterium glutamicum ATCC 13032]
HG1000786N0_160000_gene_prediction1	263	0.06	gi|5922612|dbj|BAA84613.1|	unnamed protein product [Oryza sativa (japonica cultivar-group)]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000799N0_20000_gene_prediction1	54	0.4	gi|21224658|ref|NP_630437.1|	putative secreted chitinase [Streptomyces coelicolor A3(2)]
HG1000824N0_160000_gene_prediction1	258	0.06	gi|9631237|ref|NP_048019.1|	orf 46 [Ateline herpesvirus 3] gi|11264502|pir||T42961 probable uracil-DNA glycosylase (EC
HG1000824N0_10000_gene_prediction1	258	0.06	gi|9631237|ref|NP_048019.1|	orf 46 [Ateline herpesvirus 3] gi|11264502|pir||T42961 probable uracil-DNA glycosylase (EC
HG1000839N0_160000_gene_prediction1	545	0.78	gi|26339496|dbj|BAC33419.1|	unnamed protein product [Mus musculus] gi|26347029|dbj|BAC37163.1| unnamed protein product [Mus musculus]
HG1000869N0_160000_gene_prediction1	548	0.31	gi|22749227|ref|NP_689810.1|	hypothetical protein FLJ35989 [Homo sapiens]
HG1000904N0_160000_gene_prediction2	484	0.69	gi|106322|pir||B34087	hypothetical protein (L1H 3' region) - human

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1000904N0_40000_gene_prediction1	420	0.95	gi|106322|pir||B34087	hypothetical protein (L1H 3'region) - human
HG1001187N0_160000_gene_prediction1	169	0.79	gi|125649|sp|P28028|KRAB_MOUSE	B-RAF proto-oncogene serine/threonine-protein kinase
HG1001192N0_160000_gene_prediction1	1292	0.75	gi|2072948|gb|AAC51261.1|	putative p150 [Homo sapiens]
HG1001199N0_160000_gene_prediction1	717	0.94	gi|19923640|ref|NP_115756.2|	mitochondrial elongation factor G2 isoform 1; elongation factor G2; MSTP027 [Homo sapiens]
HG1001199N0_160000_gene_prediction2	717	0.94	gi|19923640|ref|NP_115756.2|	mitochondrial elongation factor G2 isoform 1; elongation factor G2; MSTP027 [Homo sapiens]
HG1001220N0_160000_gene_prediction1	275	0.26	gi|27485433|ref|XP_208200.1|	similar to helix-destabilizing protein - rat [Homo sapiens]
HG1001229N0_160000_gene_prediction2	217	0.08	gi|29124855|gb|AAO63577.1|	secreted protein 5 precursor [Ancylostoma caninum]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1001230N0_5000_gene_prediction1	94	0.24	gi|28436723|gb|AAH47067.1|	Similar to RIKEN cDNA 5730521P14 gene [Mus musculus]
HG1001235N0_160000_gene_prediction1	70	0.22	gi|20089534|ref|NP_615609.1|	predicted protein [Methanosarcina acetivorans str. C2A]
HG1001235N0_10000_gene_prediction1	457	0.5	gi|27477710|ref|XP_098238.6|	similar to hypothetical protein DKFZp434D0215.1 - human (fragment) [Homo sapiens]
HG1001235N0_20000_gene_prediction1	529	0.57	gi|27477710|ref|XP_098238.6|	similar to hypothetical protein DKFZp434D0215.1 - human (fragment) [Homo sapiens]
HG1001235N0_160000_gene_prediction2	1020	0.75	gi|27477710|ref|XP_098238.6|	similar to hypothetical protein DKFZp434D0215.1 - human (fragment) [Homo sapiens]
HG1001235N0_160000_gene_prediction3	404	0.12	gi|4884836|gb|AAD31829.1|	NapG oxidoreductase [Streptomyces collinus]
HG1001260N0_160000_gene_prediction1	121	0.24	gi|23016062|gb|ZP_00055822.1|	hypothetical protein [Magnetospirillum magnetotacticum]
HG1001260N0_40000_gene_prediction1	172	0.17	gi|23016062|gb|ZP_00055822.1|	hypothetical protein [Magnetospirillum magnetotacticum]
HG1001264N0_160000_gene_prediction1	502	0.66	gi|23273545|gb|AAH35937.1|	Unknown (protein for MGC:25996) [Homo sapiens]

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1001274N0_160000_gene_prediction1	479	0.63	gi|17980446|gb|AAL50636.1|	unknown [Homo sapiens]
HG1001313N0_160000_gene_prediction1	240	0.09	gi|22326598|ref|NP_196037.2|	EF - hand Calcium binding protein - like; protein id: At5g04170.1, supported by cDNA:
HG1001335N0_160000_gene_prediction1	225	0.58	gi|25024189|ref|XP_204250.1|	similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1001335N0_160000_gene_prediction2	225	0.58	gi|25024189|ref|XP_204250.1|	similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1001417N0_160000_gene_prediction1	860	0.18	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4 regulatory subunit 2 [Homo sapiens]
HG1001417N0_1000_gene_prediction1	264	0.62	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4 regulatory subunit 2 [Homo sapiens]
HG1001417N0_160000_gene_prediction2	771	0.21	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4 regulatory subunit 2 [Homo sapiens]
HG1001417N0_160000_gene_prediction3	1083	0.14	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4 regulatory subunit 2 [Homo sapiens]
HG1001439N0_160000_gene_prediction1	99	0.23	gi|7439877|pir||JC5932	high mobility group I HMGI chromosomal protein isoform C-beta - human

FP ID	Length, Predicted Protein	Prediction Covered by Public	Top Hit Accession No.	Top Hit Annotation
HG1001485N0_10000_gene_prediction1	478	0.43	gi|106322|pir||B34087	hypothetical protein (L1H 3' region) - human

Table 2. Characteristics of the Claimed Sequences, and of the Human Protein With the Highest Degree of Similarity to Each

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000569N0_160000_gene_prediction1	1739	no_pfam	0.99	(3-21)		0	gi|25090171|sp|Q8TEK3|DOTL_HUMAN	Histone-lysine N-methyltransferase,
HG1001052N0_0_gene_prediction1	566	no_pfam	0.94	(1-31)		0	gi|11038659|ref|NP_067610.1|	a disintegrin and metalloprotease
HG1000498N0_160000_gene_prediction1		no_pfam	0.9	(1-20)		0		no_human_hit
HG1000685N0_160000_gene_prediction1	1091	no_pfam	0.87	(1-25)		0	gi|22770596|gb|AAN06673.1|	voltage-gated calcium channel
HG1000622N0_160000_gene_prediction1	712	no_pfam	0.52	(1-30)		0	gi|1335205|emb|CAA36480.1|	ORFII [Homo sapiens]
HG1000390N0_1000_gene_prediction1	950	no_pfam	0.51	(2-18)		0	gi|22044951|ref|XP_087331.5|	similar to URB [Homo sapiens]
HG1000806N0_20000 gene prediction	783	no_pfam	0.43	(15-33)		0	gi|23943866|ref|NP_066988.1|	hypothetical protein from clone

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
1							6988.1|	clone
HG1001489N0_20000_gene_prediction1	783	no_pfam	0.43	(15-33)		0	gi|23943866|ref|NP_066988.1|	hypothetical protein from clone
HG1001478N0_10000_gene_prediction1		no_pfam	0.38	(1-16)		0		no_human_hit
HG1000806N0_160000_gene_prediction1	783	no_pfam	0.37	(15-33)		0	gi|23943866|ref|NP_066988.1|	hypothetical protein from clone
HG1000403N0_160000_gene_prediction1		no_pfam	0.32	(2-15)		0		no_human_hit
HG1001201N0_160000_gene_prediction1	501	no_pfam	0.3	(1-31)		0	gi|27552814|gb|AAH41160.1|	Similar to hypothetical protein LOC55580
HG1000617N0_20000_gene_prediction1	1644	no_pfam	0.22	(15-47)		0	gi|4240137|dbj|BAA74847.1|	KIAA0824 protein [Homo sapiens]
HG1001334N0_160000_gene_predictio		no_pfam	0.22	(1-24)		0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
n1								
HG1000834N0_160000_gene_prediction1	338	Transposase_22	0.19	(12-27)		0	gi|2072955|gb|AAC51265.1|	p40 [Homo sapiens]
HG1000752N0_10000_gene_prediction1		no_pfam	0.17	(3-21)		0		no_human_hit
HG1000839N0_160000_gene_prediction2	412	no_pfam	0.17	(1-22)	(161-183)(322-344)(359-381)(388-407)	4	gi|19923957|ref|NP_612440.1|	hypothetical protein BC011982 [Homo
HG1000360N0_20000_gene_prediction1	790	no_pfam	0.16	(1-20)		0	gi|25071868|ref|XP_147545.2|	similar to DKFZP586M1824 protein [Homo
HG1000360N0_20000_gene_prediction2	790	no_pfam	0.15	(1-20)		0	gi|25071868|ref|XP_147545.2|	similar to DKFZP586M1824 protein [Homo
HG1000559N0_10000_gene_prediction	149	no_pfam	0.12	(19-32)		0	gi|18557931|ref|XP_087525.1|	similar to PRO1546 [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
1							7525.1|	sapiens]
HG1000570N0_160000_gene_prediction1		no_pfam	0.11	(7-25)		0		no_human_hit
HG1000617N0_40000_gene_prediction1	1275	rvt	0.11	(1-24)		0	gi|2072951|gb|AAC51263.1|	putative p150 [Homo sapiens]
HG1000615N0_160000_gene_prediction2	173	KOW	0.1	(1-13)		0	gi|22061362|ref|XP_171933.1|	similar to ribosomal protein S4,
HG1000617N0_160000_gene_prediction1	1275	rvt	0.1	(1-24)		0	gi|2072951|gb|AAC51263.1|	putative p150 [Homo sapiens]
HG1000621N0_160000_gene_prediction2	173	KOW	0.1	(1-13)		0	gi|22061362|ref|XP_171933.1|	similar to ribosomal protein S4,
HG1001502N0_160000_gene_prediction2	290	NAP	0.09	(1-30)		0	gi|2136258|pir||I59377	template activating factor-I, splice form
HG1000501N0_160000_gene_prediction1	1198	no_pfam	0.08	(19-49)		0	gi|27720977|ref|XP_236481.1|	similar to KIAA1417 protein [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
n1								
HG1000656N0_10000_gene_prediction1		no_pfam	0.08	(14-37)		0		no_human_hit
HG1000656N0_10000_gene_prediction2		no_pfam	0.08	(14-37)		0		no_human_hit
HG1000750N0_160000_gene_prediction1	693	no_pfam	0.08	(1-19)		0	gi|8923699|ref|NP_061122.1|	golgin-like protein [Homo sapiens]
HG1001012N0_160000_gene_prediction1	163	no_pfam	0.08	(15-46)		0	gi|22325370|ref|NP_680544.1|	MGC29635 protein [Homo sapiens]
HG1001237N0_10000_gene_prediction1		no_pfam	0.08	(1-15)		0		no_human_hit
HG1000390N0_160000_gene_prediction1	950	rvt	0.07	(1-19)		0	gi|22044951|ref|XP_087331.5|	similar to URB [Homo sapiens]
HG1000847N0_10000_gene_prediction		no_pfam	0.07	(13-27)		0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
1								
HG1000331N0_160000_gene_prediction1	281	LRRCT LRRNT	0.05	(1-18)	(242-264)	1	gi|27499231|ref|XP_166856.3|	similar to SLIT1-Sa splicing product
HG1000391N0_160000_gene_prediction2	477	no_pfam	0.05	(12-26)		0	gi|4885239|ref|NP_005240.1|	forkhead box G1B; forkhead
HG1000597N0_160000_gene_prediction1	1047	no_pfam	0.04			0	gi|28488582|ref|XP_284442.1|	similar to chromosome 15 open reading
HG1000415N0_10000_gene_prediction1	467	no_pfam	0.04	(6-37)		0	gi|13959539|sp|O15198|SMA9_HUMAN	Mothers against decapentaplegic
HG1000618N0_10000_gene_prediction1		no_pfam	0.04	(1-30)		0		no_human_hit
HG1001197N0_160000_gene_prediction1	96	no_pfam	0.04	(8-38)		0	gi|27498170|ref|XP_209693.1|	similar to hypothetical protein
HG1000599N0_5000		no_pfam	0.03	(1-15)		0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
gene_prediction1								
HG1000424N0_5000_gene_prediction1	187	no_pfam	0.03	(1-15)		0	gi|27477816|ref|XP_209654.1|	similar to hypothetical protein (L1H 3
HG1001485N0_5000_gene_prediction1	1380	ig	0.03	(1-26)		0	gi|10047201|dbj|BAB13394.1|	KIAA1568 protein [Homo sapiens]
HG1000674N0_160000_gene_prediction1		no_pfam	0.02	(9-21)		0		no_human_hit
HG1000339N0_160000_gene_prediction1	95	no_pfam	0.02			0	gi|18603697|ref|XP_085597.1|	hypothetical protein XP_085597 [Homo
HG1000340N0_160000_gene_prediction1	432	LUC7	0.02			0	gi|19923485|ref|NP_057508.2|	cisplatin resistance-associated
HG1000344N0_160000_gene_prediction1		no_pfam	0.02	(18-40)		0		no_human_hit
HG1000365N0_20000_gene_prediction1	128	histone	0.02	(1-21)		0	gi|4504255|ref|NP_002097.1|	H2A histone family, member Z; H2AZ

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
1								
HG1000384N0_160000_gene_prediction1	784	no_pfam	0.02			0	gi|6624128|gb|AAF19255.1|AC004858_3 U1	small ribonucleoprotein
HG1000448N0_160000_gene_prediction1	993	no_pfam	0.02	(1-17)		0	gi|18087335|gb|AAL58838.1|AF390028_1	serine/threonine protein
HG1000506N0_160000_gene_prediction1	784	no_pfam	0.02			0	gi|6624128|gb|AAF19255.1|AC004858_3	U1 small ribonucleoprotein
HG1000550N0_160000_gene_prediction1	784	no_pfam	0.02			0	gi|6624128|gb|AAF19255.1|AC004858_3	U1 small ribonucleoprotein
HG1000647N0_160000_gene_prediction1	246	no_pfam	0.02	(1-19)		0	gi|5031905|ref|NP_005577.1|	MyoD family inhibitor [Homo sapiens]
HG1000688N0_160000_gene_prediction1	1205	Pep_M12B_propep	0.02	(1-20)		0	gi|21903371|sp|O15072|ATS3_HUMAN	ADAMTS-3 precursor (A disintegrin

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000904N0_160000_gene_prediction1	1275	rvt	0.02			0	gi|2072948|gb|AAC51261.1|	putative p150 [Homo sapiens]
HG1001417N0_5000_ gene_prediction1	417	no_pfam	0.02			0	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4
HG1001485N0_160000_gene_prediction1	1380	ig	0.02	(1-26)		0	gi|10047201|dbj|BAB13394.1|	KIAA1568 protein [Homo sapiens]
HG1001502N0_160000_gene_prediction1	1011	Peptidase_M1	0.01		(98-117)	1	gi|2209278|gb|AAB66673.1|	oxytocinase splice variant 2 [Homo sapiens]
HG1000343N0_160000_gene_prediction1	473	rvt	0.01			0	gi|21750183|dbj|BAC03736.1|	unnamed protein product [Homo sapiens]
HG1000343N0_160000_gene_prediction2		no_pfam	0.01			0		no_human_hit
HG1000369N0_160000_gene_predictio		no_pfam	0.01			0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
n1								
HG1000378N0_160000_gene_prediction1	1117	Gag_p24 dUTPase	0.01			0	gi|5802824|gb|AAD51799.1|AF164615_1	Gag-Pro-Pol protein [Homo
HG1000387N0_160000_gene_prediction1	1259	no_pfam	0.01			0	gi|126295|sp|P08547|LIN1_HUMAN	LINE-1 REVERSE TRANSCRIPTASE HOMOLOG
HG1000387N0_160000_gene_prediction2	149	no_pfam	0.01			0	gi|18557931|ref|XP_087525.1|	similar to PRO1546 [Homo sapiens]
HG1000408N0_160000_gene_prediction1		no_pfam	0.01	(1-21)		0		no_human_hit
HG1000431N0_160000_gene_prediction1	147	no_pfam	0.01			0	gi|27482385|ref|XP_208702.1|	hypothetical protein XP_208702 [Homo
HG1000457N0_160000_gene_prediction1	583	no_pfam	0.01			0	gi|13325136|gb|AAH04381.1|AAH04381	Similar to liver-specific
HG1000458N0_160000_gene_prediction1	133	no_pfam	0.01		(22-44)	1	gi|27704378|ref|XP_230791.1|	similar to mannosidase, beta A,

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
n1								
HG1000463N0_160000_gene_prediction1	166	no_pfam	0.01			0	gi|5031635|ref|NP_005498.1|	cofilin 1 (non-muscle) [Homo sapiens]
HG1000463N0_160000_gene_prediction2	2121	no_pfam	0.01	(1-18)		0	gi|27707612|ref|XP_231287.1|	similar to MLL5 [Homo sapiens] [Rattus
HG1000481N0_160000_gene_prediction1	130	no_pfam	0.01	(17-30)		0	gi|17436547|ref|XP_067994.1|	similar to heat shock factor binding
HG1000592N0_160000_gene_prediction1	496	no_pfam	0.01		(86-108)(133-155)(212-234)(238-260)(591-613)	5	gi|14017873|dbj|BAB47457.1|	KIAA1828 protein [Homo sapiens]
HG1000608N0_160000_gene_prediction1	42	no_pfam	0.01			0	gi|14249206|ref|NP_116044.1|	hypothetical protein MGC10997 [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000615N0_160000_gene_prediction1		no_pfam	0.01			0		no_human_hit
HG1000621N0_160000_gene_prediction1	1259	no_pfam	0.01			0	gi|126295|sp|P08547|LIN1_HUMAN	LINE-1 REVERSE TRANSCRIPTASE HOMOLOG
HG1000621N0_160000_gene_prediction3	149	no_pfam	0.01			0	gi|18557931|ref|XP_087525.1|	similar to PRO1546 [Homo sapiens]
HG1000652N0_160000_gene_prediction1	338	Transposase_22	0.01	(1-24)		0	gi|2072971|gb|AAC51275.1|	p40 [Homo sapiens]
HG1000663N0_160000_gene_prediction1	463	no_pfam	0.01			0	gi|9392624|gb|AAF87223.1|AF247180_1	sialic acid-binding
HG1000700N0_160000_gene_prediction1	136	no_pfam	0.01	(11-33)		0	gi|27479609|ref|XP_209735.1|	similar to LINE-1 REVERSE
HG1000701N0_160000_gene_prediction1	1608	no_pfam	0.01			0	gi|16507208|ref|NP_055940.2|	capicua (Drosophila) homolog [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000709N0_160000_gene_prediction1		no_pfam	0.01			0		no_human_hit
HG1000720N0_160000_gene_prediction1	2472	no_pfam	0.01			0	gi|7630181|dbj|BAA94765.1|	alpha1A-voltage-dependent calcium channel
HG1000786N0_160000_gene_prediction2	673	no_pfam	0.01			0	gi|7662176|ref|NP_055546.1|	ProSAPiP1 protein [Homo sapiens]
HG1000848N0_160000_gene_prediction1	471	no_pfam	0.01			0	gi|28277065|gb|AAH45626.1|	Similar to Pbx/knotted 1 homeobox 2
HG1000904N0_160000_gene_prediction3		no_pfam	0.01			0		no_human_hit
HG1001214N0_200000_gene_prediction1		no_pfam	0.01			0		no_human_hit
HG1001229N0_160000_gene_prediction1	1275	rvt	0.01			0	gi|2072953|gb|AAC51264.1|	putative p150 [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001468N0_160000_gene_prediction1	199	no_pfam	0.01			0	gi|21749211|dbj|BAC03554.1|	unnamed protein product [Homo sapiens]
HG1000382N0_160000_gene_prediction1	1613	EGF ldl_recept_a ldl_recept_b	0		(1367-1389)	1	gi|4505017|ref|NP_002327.1|	low density lipoprotein receptor-related
HG1000591N0_160000_gene_prediction1	2551	EGF Fasciclin Xlink	0		(1805-1827)	1	gi|22779441|dbj|BAC15608.1|	FELE-2 [Homo sapiens]
HG1000904N0_160000_gene_prediction4	1033	sushi	0		(1008-1027)(1034-1056)	2	gi|117315|sp|P20023|CR2_HUMAN	Complement receptor type 2 precursor
HG1000267N0_5000_gene_prediction1	335	gpdh gpdh_C	0			0	gi|7669492|ref|NP_002037.2|	glyceraldehyde-3-phosphate
HG1000332N0_10000_gene_prediction1	154	no_pfam	0			0	gi|16550911|dbj|BAB71077.1|	unnamed protein product [Homo

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000341N0_5000_gene_prediction1	1581	no_pfam	0			0	gi|28559039|ref|NP_004765.2|.	peroxisome proliferator-activated
HG1000341N0_10000_gene_prediction1	256	no_pfam	0			0	gi|16552274|dbj|BAB71280.1|	unnamed protein product [Homo sapiens]
HG1000353N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1000359N0_160000_gene_prediction1	1275	rvt Transposase_22	0			0	gi|2072961|gb|AAC51269.1|	putative p150 [Homo sapiens]
HG1000363N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1000364N0_160000_gene_prediction1	1284	no_pfam	0			0	gi|21426922|gb|AAC17708.2|	PELP1 [Homo sapiens]
HG1000367N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000379N0_160000_gene_prediction1	1275	rvt	0			0	gi|2072951|gb|AAC51263.1|	putative p150 [Homo sapiens]
HG1000390N0_10000_gene_prediction1	950	no_pfam	0			0	gi|22044951|ref|XP_087331.5|	similar to URB [Homo sapiens]
HG1000390N0_5000_gene_prediction1	950	no_pfam	0			0	gi|22044951|ref|XP_087331.5|	similar to URB [Homo sapiens]
HG1000391N0_160000_gene_prediction1	1224	no_pfam	0			0	gi|22538401|ref|NP_055884.1|	OLF-1/EBF associated zinc finger
HG1000407N0_160000_gene_prediction1		Transposase_8 rve	0	(19-42)		0		no_human_hit
HG1000408N0_160000_gene_prediction2		no_pfam	0			0		no_human_hit
HG1000429N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000431N0_20000_gene_prediction1	222	no_pfam	0			0	gi|10436424|dbj|BAB14833.1|	unnamed protein product [Homo sapiens]
HG1000435N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1000474N0_5000_gene_prediction1	964	no_pfam	0	(1-19)		0	gi|14755300|ref|XP_042833.1|	similar to KIAA0295 [Homo sapiens]
HG1000489N0_160000_gene_prediction1	1275	rvt Exo_endo_phos	0	(1-24)		0	gi|2136112|pir||S65824	reverse transcriptase homolog - human
HG1000499N0_160000_gene_prediction1	455	no_pfam	0			0	gi|21757162|dbj|BAC05041.1|	unnamed protein product [Homo sapiens]
HG1000500N0_160000_gene_prediction1	281	7tm_1	0		(25-47)(60-82)(97-119)(201-223)	4	gi|27482474|ref|XP_068681.3|	similar to seven transmembrane helix

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000505N0_160000_gene_prediction1	287	no_pfam	0			0	gi|27730259|ref|XP_217905.1|	similar to RNA-binding protein S1,
HG1000509N0_10000_gene_prediction1	964	no_pfam	0	(1-19)		0	gi|14755300|ref|XP_042833.1|	similar to KIAA0295 [Homo sapiens]
HG1000519N0_160000_gene_prediction1	2285	no_pfam	0			0	gi|21264565|ref|NP_006006.3|	SWI/SNF-related matrix-associated
HG1000521N0_160000_gene_prediction1	5179	no_pfam	0			0	gi|4505285|ref|NP_002448.1|	mucin 2 [Homo sapiens]
HG1000549N0_160000_gene_prediction1	1275	rvt	0			0	gi|2072977|gb|AAC51279.1|	putative p150 [Homo sapiens]
HG1000549N0_160000_gene_prediction2	630	no_pfam	0			0	gi|15826862|ref|NP_296375.1|	JM11 protein [Homo sapiens]
HG1000549N0_160000_gene_prediction3	1275	rvt	0			0	gi|2072977|gb|AAC51279.1|	putative p150 [Homo sapiens]

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000562N0_160000_gene_prediction1	557	no_pfam	0			0	gi|4506089|ref|NP_002738.1|	mitogen-activated protein kinase 4;
HG1000582N0_160000_gene_prediction1	729	Neurochondrin	0			0	gi|24307949|ref|NP_055099.1|	neurochondrin [Homo sapiens]
HG1000598N0_160000_gene_prediction1	688	zf-C2H2 BTB	0			0	gi|4885419|ref|NP_005332.1|	GLI-Kruppel family member HKR3 [Homo
HG1000606N0_20000_gene_prediction1	445	no_pfam	0			0	gi|14270361|emb|CAC39434.1|	histamine H3 receptor [Homo sapiens]
HG1000607N0_160000_gene_prediction1	1275	rvt Exo_endo_phos	0			0	gi|339777|gb|AAB59368.1|	ORF2 contains a reverse transcriptase
HG1000608N0_20000_gene_prediction1	42	no_pfam	0			0	gi|14249206|ref|NP_116044.1|	hypothetical protein MGC10997 [Homo
HG1000616N0_1000_gene_prediction1		no_pfam	0	(1-18)		0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000622N0_160000_gene_prediction2		no_pfam	0	(1-34)		0		no_human_hit
HG1000625N0_160000_gene_prediction1	1275	Exo_endo_phos	0			0	gi|2072967|gb|AAC51273.1|	putative p150 [Homo sapiens]
HG1000628N0_40000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1000628N0_20000_gene_prediction1	1275	rvt	0			0	gi|2072972|gb|AAC51276.1|	putative p150 [Homo sapiens]
HG1000638N0_5000_gene_prediction1		no_pfam	0	(1-18)		0		no_human_hit
HG1000642N0_160000_gene_prediction1	1275	rvt Exo_endo_phos	0			0	gi|5070622|gb|AAD39215.1|AF148856_2	unknown [Homo sapiens]
HG1000646N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000649N0_160000_gene_prediction1	420	rvt	0			0	gi|11037117|gb|AAG27485.1|AF194537_1	NAG13 [Homo sapiens]
HG1000652N0_160000_gene_prediction2	145	no_pfam	0	(1-19)		0	gi|27498807|ref|XP_208381.1|	similar to nuclear receptor binding
HG1000656N0_160000_gene_prediction1	630	no_pfam	0			0	gi|7513073|pir||T00351	hypothetical protein KIAA0707 - human
HG1000656N0_160000_gene_prediction2	630	no_pfam	0			0	gi|7513073|pir||T00351	hypothetical protein KIAA0707 - human
HG1000670N0_160000_gene_prediction1	1208	DEAD	0			0	gi|4759030|ref|NP_004251.1|	RecQ protein-like 4; RecQ protein 4 [Homo
HG1000685N0_160000_gene_prediction2		no_pfam	0	(1-17)		0		no_human_hit
HG1000700N0_160000_gene_prediction2	598	no_pfam	0			0	gi|22538495|ref|NP_071924.1|	Williams Beuren syndrome chromosome

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000738N0_160000_gene_prediction1	824	no_pfam	0			0	gi|6912492|ref|NP_036456.1|	mitogen-activated protein kinase 8
HG1000739N0_160000_gene_prediction1	338	Transposase_22	0			0	gi|5070621|gb|AAD39214.1|AF148856_1	unknown [Homo sapiens]
HG1000739N0_160000_gene_prediction2	338	Transposase_22	0			0	gi|17980446|gb|AAL50636.1|	unknown [Homo sapiens]
HG1000779N0_160000_gene_prediction1	1741	no_pfam	0			0	gi|14577919|ref|NP_009224.1|	complement component 4A preproprotein;
HG1000786N0_160000_gene_prediction1	124	no_pfam	0			0	gi|2565091|gb|AAB91452.1|	CTG26 alternate open reading frame [Homo
HG1000799N0_20000_gene_prediction1	891	no_pfam	0	(1-8)		0	gi|23488603|gb|EAA21358.1|	Homo sapiens dJ846D11.1, putative
HG1000824N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1000824N0_10000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1000839N0_160000_gene_prediction1	640	LRRCT ig	0	(8-23)	(436-458)	1	gi|20560455|ref|XP_045271.5|	similar to NAG14 protein [Homo
HG1000869N0_160000_gene_prediction1	581	no_pfam	0			0	gi|22749227|ref|NP_689810.1|	hypothetical protein FLJ35989 [Homo
HG1000904N0_160000_gene_prediction2	1275	rvt	0			0	gi|2072953|gb|AAC51264.1|	putative p150 [Homo sapiens]
HG1000904N0_40000_gene_prediction1	1275	rvt	0			0	gi|2072977|gb|AAC51279.1|	putative p150 [Homo sapiens]
HG1001187N0_160000_gene_prediction1	765	pkinase	0			0	gi|4757868|ref|NP_004324.1|	v-raf murine sarcoma viral oncogene

WO 2004/020595 PCT/US2003/027107

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
n1								
HG1001192N0_160000_gene_predictio n1	1275	rvt Transposase_22	0			0	gi|2072948|gb|AAC51261.1|	putative p150 [Homo sapiens]
HG1001199N0_160000_gene_predictio n1	779	EFG_C EFG_IV GTP_EFTU_D2 GTP_EFTU	0			0	gi|19923640|ref|NP_115756.2|	mitochondrial elongation factor G2
HG1001199N0_160000_gene_predictio n2	779	EFG_C EFG_IV GTP_EFTU_D2 GTP_EFTU	0			0	gi|19923640|ref|NP_115756.2|	mitochondrial elongation factor G2
HG1001220N0_160000_gene_predictio n1	321	rrm	0	(15-44)		0	gi|27485433|ref|XP_208200.1|	similar to helix-destabilizing protein
HG1001229N0_160000_gene_predictio n2		no_pfam	0			0		no_human_hit

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001230N0_5000_gene_prediction1	546	no_pfam	0	(1-23)		0	gi|7022205|dbj|BAA91516.1|	unnamed protein product [Homo sapiens]
HG1001235N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1001235N0_10000_gene_prediction1	767	no_pfam	0			0	gi|27477710|ref|XP_098238.6|	similar to hypothetical protein
HG1001235N0_20000_gene_prediction1	767	no_pfam	0			0	gi|27477710|ref|XP_098238.6|	similar to hypothetical protein
HG1001235N0_160000_gene_prediction2	767	SH3	0			0	gi|27477710|ref|XP_098238.6|	similar to hypothetical protein
HG1001235N0_160000_gene_prediction3	1004	no_pfam	0			0	gi|24429582|ref|NP_722520.1| f	riend of GATA-1; zinc finger protein
HG1001260N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit

WO 2004/020595 PCT/US2003/027107

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001260N0_40000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1001264N0_160000_gene_prediction1	542	no_pfam	0			0	gi|23273545|gb|AAH35937.1|	Unknown (protein for MGC:25996) [Homo
HG1001274N0_160000_gene_prediction1	338	Transposase_22	0			0	gi|17980446|gb|AAL50636.1|	unknown [Homo sapiens]
HG1001313N0_160000_gene_prediction1	791	no_pfam	0			0	gi|26251997|gb|AAH40541.1|	Unknown (protein for IMAGE:5267781)
HG1001335N0_160000_gene_prediction1	1275	rvt	0			0	gi|2072977|gb|AAC51279.1|	putative p150 [Homo sapiens]
HG1001335N0_160000_gene_prediction2	1275	rvt	0			0	gi|2072977|gb|AAC51279.1|	putative p150 [Homo sapiens]
HG1001417N0_160000_gene_prediction1	417	no_pfam	0			0	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4

FP ID	Length, Human Top Hit	Pfam	Tree Vote, Secreted	SP Positions	TM Domains	TM Total	Top Human Hit Accession No.	Top Human Hit Annotation
HG1001417N0_1000_gene_prediction1	417	no_pfam	0			0	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4
HG1001417N0_160000_gene_prediction2	417	no_pfam	0			0	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4
HG1001417N0_160000_gene_prediction3	417	no_pfam	0			0	gi|22043508|ref|XP_041350.5|	similar to protein phosphatase 4
HG1001439N0_160000_gene_prediction1		no_pfam	0			0		no_human_hit
HG1001485N0_10000_gene_prediction1	1259	rvt Exo_endo_phos	0			0	gi|126295|sp|P08547|LIN1_HUMAN	LINE-1 REVERSE TRANSCRIPTASE HOMOLOG

Table 3. Characteristics of the Fantom Mouse Protein With the Highest Degree of Similarity to the Claimed Sequences

FP ID	Fantom Top Hit Annotation
HG1000214N0_160000_gene_prediction1	pre-B lymphocyte gene 1 [Mus musculus]
HG1000323N0_160000_gene_prediction1	lipoprotein lipase [Mus musculus]
HG1000323N0_160000_gene_prediction2	similar to procollagen, type V, alpha 2 [Mus musculus]
HG1000327N0_1000_gene_prediction1	unnamed protein product [Mus musculus]
HG1000327N0_160000_gene_prediction1	unnamed protein product [Mus musculus]
HG1000434N0_160000_gene_prediction1	uromodulin; Tamm-Horsfall glycoprotein [Mus musculus]
HG1000449N0_160000_gene_prediction1	trefoil factor 1 [Mus musculus]
HG1000807N0_160000_gene_prediction1	IGFBP-like protein [Mus musculus].
HG1000807N0_5000_gene_prediction1	gi|9055246|ref|NP_061211.1| IGFBP-like protein [Mus musculus]
HG1001280N0_160000_gene_prediction1	gi|26336763|dbj|BAC32064.1| unnamed protein product [Mus musculus]
HG1000193N0_160000_gene_prediction1	gi|21595011|gb|AAH31409.1| RIKEN cDNA 2410030O07 gene [Mus musculus]
HG1000286N0_160000_gene_prediction1	gi|303678|dbj|BAA02298.1| 47-kDa heat shock protein [Mus musculus]
HG1000569N0_160000_gene_prediction1	gi|20881983|ref|XP_122793.1| similar to heat-stable antigen-related hypothetical protein HSA-C - mouse [Mus musculus]
HG1000992N0_160000_gene_prediction1	gi|26331916|dbj|BAC29688.1| unnamed protein product [Mus musculus]
HG1001148N0_160000_gene_prediction1	gi|6752962|ref|NP_033744.1| a disintegrin and metalloprotease domain 15 (metargidin); a disintegrin and metalloproteinase domain (ADAM) 15 (metargidin) [Mus musculus]
HG1001185N0_160000_gene_prediction2	gi|26329785|dbj|BAC28631.1| unnamed protein product [Mus musculus]
HG1001280N0_5000_gene_prediction1	gi|26336763|dbj|BAC32064.1| unnamed protein product [Mus musculus]
HG1001302N0_160000_gene_prediction2	gi|20136122|gb|AAM11539.1| matrilin-2 [Mus musculus]
HG1000361N0_160000_gene_prediction1	gi|20867549|ref|XP_125932.1| RIKEN cDNA 9030421L11 [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000361N0_20000_gene_prediction 1	gi|26330472|dbj|BAC28966.1| unnamed protein product [Mus musculus]
HG1000792N0_160000_gene_predictio n1	gi|27229118|ref|NP_082129.2| RIKEN cDNA 0610006F02 [Mus musculus]
HG1000934N0_160000_gene_predictio n1	gi|20867549|ref|XP_125932.1| RIKEN cDNA 9030421L11 [Mus musculus]
HG1000976N0_160000_gene_predictio n1	gi|11967965|ref|NP_071879.1| cytochrome P450, subfamily IVF, polypeptide 14 (leukotriene B4 omega hydroxylase) [Mus musculus]
HG1000992N0_10000_gene_prediction 1	gi|26331916|dbj|BAC29688.1| unnamed protein product [Mus musculus]
HG1001185N0_1000_gene_prediction1	gi|26329785|dbj|BAC28631.1| unnamed protein product [Mus musculus]
HG1001185N0_160000_gene_predictio n1	gi|26329785|dbj|BAC28631.1| unnamed protein product [Mus musculus]
HG1001185N0_1000_gene_prediction2	gi|26329785|dbj|BAC28631.1| unnamed protein product [Mus musculus]
HG1001185N0_5000_gene_prediction1	gi|26329785|dbj|BAC28631.1| unnamed protein product [Mus musculus]
HG1001280N0_10000_gene_prediction 1	gi|26336763|dbj|BAC32064.1| unnamed protein product [Mus musculus]
HG1000361N0_10000_gene_prediction 1	gi|26330472|dbj|BAC28966.1| unnamed protein product [Mus musculus]
HG1001381N0_1000_gene_prediction1	gi|26343077|dbj|BAC35195.1| unnamed protein product [Mus musculus]
HG1000263N0_5000_gene_prediction1	gi|26360198|dbj|BAB25612.2| unnamed protein product [Mus musculus]
HG1001052N0_0_gene_prediction1	gi|20072693|gb|AAH27297.1| Similar to cyclin K [Mus musculus]
HG1000498N0_160000_gene_predictio n1	gi|26352844|dbj|BAC40052.1| unnamed protein product [Mus musculus]
HG1000579N0_160000_gene_predictio n1	gi|26330550|dbj|BAC29005.1| unnamed protein product [Mus musculus]
HG1000685N0_160000_gene_predictio n1	gi|6753236|ref|NP_033915.1| calcium channel, voltage dependent, alpha2/delta subunit 3; alpha 2 delta-3 [Mus musculus]
HG1000191N0_160000_gene_predictio n1	gi|13385832|ref|NP_080608.1| RIKEN cDNA 1810055D05 [Mus musculus]
HG1000296N0_160000_gene_predictio n2	gi|25054735|ref|XP_192839.1| ATPas, class II, type 9B [Mus musculus]
HG1000346N0_1000_gene_prediction1	gi|26330504|dbj|BAC28982.1| unnamed protein product [Mus musculus]
HG1000963N0_5000_gene_prediction1	gi|12963665|ref|NP_075892.1| mesoderm

FP ID	Fantom Top Hit Annotation
	development candiate 2; RIKEN cDNA 2210015O11 gene [Mus musculus]
HG1000610N0_160000_gene_prediction1	gi|26335037|dbj|BAC31219.1| unnamed protein product [Mus musculus]
HG1000342N0_160000_gene_prediction1	gi|20881983|ref|XP_122793.1| similar to heat-stable antigen-related hypothetical protein HSA-C - mouse [Mus musculus]
HG1000342N0_160000_gene_prediction2	gi|20881983|ref|XP_122793.1| similar to heat-stable antigen-related hypothetical protein HSA-C - mouse [Mus musculus]
HG1000650N0_20000_gene_prediction1	gi|20270210|ref|NP_083847.1| RIKEN cDNA 1110001A12 [Mus musculus]
HG1000191N0_160000_gene_prediction2	gi|13385832|ref|NP_080608.1| RIKEN cDNA 1810055D05 [Mus musculus]
HG1000449N0_160000_gene_prediction3	gi|6755773|ref|NP_035705.1| trefoil factor 3, intestinal [Mus musculus]
HG1000181N0_20000_gene_prediction1	gi|26334755|dbj|BAC31078.1| unnamed protein product [Mus musculus]
HG1001058N0_160000_gene_prediction1	gi|20344262|ref|XP_110959.1| similar to LD31582p [Drosophila melanogaster] [Mus musculus]
HG1000187N0_160000_gene_prediction2	gi|26346705|dbj|BAC37001.1| unnamed protein product [Mus musculus]
HG1000191N0_1000_gene_prediction1	gi|13385832|ref|NP_080608.1| RIKEN cDNA 1810055D05 [Mus musculus]
HG1000319N0_160000_gene_prediction1	gi|25021456|ref|XP_207950.1| similar to pORF2 [Mus musculus domesticus]
HG1000137N0_0_gene_prediction1	gi|20843789|ref|XP_133814.1| similar to hypothetical protein IMAGE3455200 [Homo sapiens] [Mus musculus]
HG1000191N0_5000_gene_prediction1	gi|12842346|dbj|BAB25565.1| unnamed protein product [Mus musculus]
HG1000622N0_160000_gene_prediction1	gi|25022040|ref|XP_204233.1| similar to ORF2 [Mus musculus domesticus]
HG1000390N0_1000_gene_prediction1	gi|20892585|ref|XP_147977.1| RIKEN cDNA 2610001E17 [Mus musculus]
HG1001350N0_5000_gene_prediction1	gi|13386102|ref|NP_080892.1| RIKEN cDNA 1500026D16 [Mus musculus]
HG1000327N0_160000_gene_prediction2	gi|26324414|dbj|BAC25961.1| unnamed protein product [Mus musculus]
HG1000179N0_160000_gene_prediction1	gi|20862121|ref|XP_146270.1| similar to putative alpha 1,3-fucosyl transferase [Mus musculus]
HG1000806N0_20000_gene_prediction	gi|23592855|ref|XP_129487.2| hypothetical

FP ID	Fantom Top Hit Annotation
1	protein MGC40674 [Mus musculus]
HG1000991N0_160000_gene_prediction1	gi|6755338|ref|NP_036013.1| ring finger protein 13 [Mus musculus]
HG1001489N0_20000_gene_prediction1	gi|23592855|ref|XP_129487.2| hypothetical protein MGC40674 [Mus musculus]
HG1001038N0_5000_gene_prediction1	gi|20892051|ref|XP_148657.1| similar to Lethal(2)neighbour of tid protein 2 (NOT53) [Mus musculus]
HG1001376N0_160000_gene_prediction2	gi|27261816|ref|NP_080861.1| RIKEN cDNA C530005J20 [Mus musculus]
HG1001376N0_20000_gene_prediction2	gi|27261816|ref|NP_080861.1| RIKEN cDNA C530005J20 [Mus musculus]
HG1001478N0_10000_gene_prediction1	gi|6979907|gb|AAF34647.1|AF221103_1 kinesin-related protein KIFC5B [Mus musculus]
HG1000806N0_160000_gene_prediction1	gi|23592855|ref|XP_129487.2| hypothetical protein MGC40674 [Mus musculus]
HG1000409N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000884N0_160000_gene_prediction1	gi|26329055|dbj|BAC28266.1| unnamed protein product [Mus musculus]
HG1000575N0_160000_gene_prediction1	gi|20889984|ref|XP_129281.1| RIKEN cDNA 4930538D17 [Mus musculus]
HG1000403N0_160000_gene_prediction1	gi|26340168|dbj|BAC33747.1| unnamed protein product [Mus musculus]
HG1000906N0_10000_gene_prediction1	gi|20836822|ref|XP_130277.1| similar to Plakophilin 4 (p0071) [Mus musculus]
HG1001201N0_160000_gene_prediction1	gi|26341746|dbj|BAC34535.1| unnamed protein product [Mus musculus]
HG1000485N0_160000_gene_prediction1	gi|23597904|ref|XP_129263.2| protein phosphatase 1, regulatory (inhibitor) subunit 3C [Mus musculus]
HG1000328N0_160000_gene_prediction1	gi|26336731|dbj|BAC32048.1| unnamed protein product [Mus musculus]
HG1000231N0_160000_gene_prediction1	gi|26341312|dbj|BAC34318.1| unnamed protein product [Mus musculus]
HG1001257N0_10000_gene_prediction1	gi|26346593|dbj|BAC36945.1| unnamed protein product [Mus musculus]
HG1000026N0_5000_gene_prediction1	gi|9506367|ref|NP_062425.1| ATP-binding cassette, sub-family B, member 10; ATP-binding cassette, sub-family B (MDR/TAP), member 12; Abc-mitochondrial erythroid [Mus musculus]
HG1000300N0_160000_gene_predictio	gi|12846244|dbj|BAB27089.1| unnamed protein

FP ID	Fantom Top Hit Annotation
n1	product [Mus musculus]
HG1000109N0_160000_gene_predictio n1	gi|22779909|ref|NP_690028.1| RIKEN cDNA 2700083B01 [Mus musculus]
HG1000617N0_20000_gene_prediction 1	gi|7949115|ref|NP_058079.1| Ser/Arg-related nuclear matrix protein; plenty-of-prolines-101; serine/arginine repetitive matrix protein 1 [Mus musculus]
HG1001110N0_160000_gene_predictio n1	gi|22779909|ref|NP_690028.1| RIKEN cDNA 2700083B01 [Mus musculus]
HG1001334N0_160000_gene_predictio n1	gi|26332062|dbj|BAC29761.1| unnamed protein product [Mus musculus]
HG1001376N0_160000_gene_predictio n3	gi|27261816|ref|NP_080861.1| RIKEN cDNA C530005J20 [Mus musculus]
HG1000026N0_20000_gene_prediction 1	gi|9506367|ref|NP_062425.1| ATP-binding cassette, sub-family B, member 10; ATP-binding cassette, sub-family B (MDR/TAP), member 12; Abc-mitochondrial erythroid [Mus musculus]
HG1000276N0_1000_gene_prediction1	gi|19527228|ref|NP_598768.1| DNA segment, Chr 10, ERATO Doi 214, expressed [Mus musculus]
HG1000822N0_160000_gene_predictio n2	gi|6680195|ref|NP_032255.1| histone deacetylase 2; DNA segment, Chr 10, Wayne State University 179, expressed [Mus musculus]
HG1000173N0_20000_gene_prediction 1	gi|26345110|dbj|BAC36204.1| unnamed protein product [Mus musculus]
HG1000834N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1001044N0_1000_gene_prediction1	gi|26330836|dbj|BAC29148.1| unnamed protein product [Mus musculus]
HG1000299N0_1000_gene_prediction1	gi|6753882|ref|NP_034349.1| FK506 binding protein 4 (59 kDa) [Mus musculus]
HG1000752N0_10000_gene_prediction 1	gi|25955698|gb|AAH40387.1| Similar to PTPL1-associated RhoGAP 1 [Mus musculus]
HG1000839N0_160000_gene_predictio n2	gi|17512422|gb|AAH19171.1| Similar to RIKEN cDNA 2310010G13 gene [Mus musculus]
HG1000659N0_160000_gene_predictio n1	gi|26333733|dbj|BAC30584.1| unnamed protein product [Mus musculus]
HG1000659N0_160000_gene_predictio n2	gi|26333733|dbj|BAC30584.1| unnamed protein product [Mus musculus]
HG1000013N0_160000_gene_predictio n1	gi|20881136|ref|XP_126284.1| similar to sperm antigen HCMOGT-1 [Homo sapiens] [Mus

FP ID	Fantom Top Hit Annotation
	musculus]
HG1000173N0_160000_gene_predictio n1	gi|26345110|dbj|BAC36204.1| unnamed protein product [Mus musculus]
HG1000330N0_160000_gene_predictio n1	gi|27462832|gb|AAO15605.1|AF462146_1 modulator of estrogen induced transcription [Mus musculus]
HG1000360N0_20000_gene_prediction 1	gi|7861746|gb|AAF70384.1|AF189263_1 GABA-A receptor epsilon-like subunit [Mus musculus]
HG1000178N0_10000_gene_prediction 1	gi|13384830|ref|NP_079706.1| RIKEN cDNA 1110066C01 [Mus musculus]
HG1000178N0_10000_gene_prediction 2	gi|13384830|ref|NP_079706.1| RIKEN cDNA 1110066C01 [Mus musculus]
HG1000360N0_20000_gene_prediction 2	gi|7861746|gb|AAF70384.1|AF189263_1 GABA-A receptor epsilon-like subunit [Mus musculus]
HG1000640N0_160000_gene_predictio n1	gi|21313034|ref|NP_080346.1| RIKEN cDNA 2900091E11 [Mus musculus]
HG1001000N0_160000_gene_predictio n1	gi|10181212|ref|NP_065613.1| RIKEN cDNA 1300007B12; clone MNCb-2755 [Mus musculus]
HG1001418N0_160000_gene_predictio n1	gi|20819462|ref|XP_158058.1| hypothetical protein XP_158058 [Mus musculus]
HG1000153N0_20000_gene_prediction 1	gi|26379523|dbj|BAB29070.2| unnamed protein product [Mus musculus]
HG1000255N0_160000_gene_predictio n1	gi|13385532|ref|NP_080303.1| RIKEN cDNA 2700086I23 [Mus musculus]
HG1000186N0_160000_gene_predictio n1	gi|20963196|ref|XP_135684.1| RIKEN cDNA 1700022L20 [Mus musculus]
HG1000259N0_160000_gene_predictio n1	gi|26360198|dbj|BAB25612.2| unnamed protein product [Mus musculus]
HG1000559N0_10000_gene_prediction 1	
HG1000084N0_10000_gene_prediction 1	gi|6678794|ref|NP_032953.1| mitogen activated protein kinase kinase 1; MAP kinase kinase 1; protein kinase, mitogen activated, kinase 1, p45 [Mus musculus]
HG1000217N0_160000_gene_predictio n1	gi|6681015|ref|NP_031789.1| cysteine rich intestinal protein [Mus musculus]
HG1000217N0_160000_gene_predictio n2	gi|6681015|ref|NP_031789.1| cysteine rich intestinal protein [Mus musculus]
HG1000329N0_160000_gene_predictio n1	gi|26330870|dbj|BAC29165.1| unnamed protein product [Mus musculus]
HG1000570N0_160000_gene_predictio	gi|6716522|gb|AAF26675.1|AF155821_1

FP ID	Fantom Top Hit Annotation
n1	CPG16 [Mus musculus]
HG1000617N0_40000_gene_prediction 1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000227N0_160000_gene_predictio n1	gi|21362402|sp|Q9CZB0|C560_MOUSE Succinate dehydrogenase cytochrome b560 subunit, mitochondrial precursor (Integral membrane protein CII-3) (QPS1) (QPs-1)
HG1000269N0_10000_gene_prediction 1	gi|7706341|ref|NP_057145.1| yippee protein [Homo sapiens]
HG1000615N0_160000_gene_predictio n2	gi|4506725|ref|NP_000998.1| ribosomal protein S4, X-linked X isoform; 40S ribosomal protein S4, X isoform; ribosomal protein S4X isoform; single-copy abundant mRNA; cell cycle gene 2 [Homo sapiens]
HG1000617N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000621N0_160000_gene_predictio n2	gi|4506725|ref|NP_000998.1| ribosomal protein S4, X-linked X isoform; 40S ribosomal protein S4, X isoform; ribosomal protein S4X isoform; single-copy abundant mRNA; cell cycle gene 2 [Homo sapiens]
HG1000990N0_160000_gene_predictio n1	gi|10946760|ref|NP_067381.1| triggering receptor expressed on myeloid cells 1; triggering receptor expressed in monocytes 1 [Mus musculus]
HG1000998N0_160000_gene_predictio n1	gi|6678483|ref|NP_033483.1| ubiquitin-activating enzyme E1, Chr X [Mus musculus]
HG1001225N0_160000_gene_predictio n1	gi|10181192|ref|NP_065589.1| sulfotransferase-related protein SULT-X1 [Mus musculus]
HG1001269N0_5000_gene_prediction1	gi|21311883|ref|NP_080887.1| RIKEN cDNA 0610007O07 [Mus musculus]
HG1001269N0_160000_gene_predictio n1	gi|21311883|ref|NP_080887.1| RIKEN cDNA 0610007O07 [Mus musculus]
HG1000103N0_160000_gene_predictio n1	gi|26327721|dbj|BAC27604.1| unnamed protein product [Mus musculus]
HG1000143N0_1000_gene_prediction1	gi|14141193|ref|NP_001004.2| ribosomal protein S9; 40S ribosomal protein S9 [Homo sapiens]
HG1000396N0_160000_gene_predictio n1	gi|25024769|ref|XP_207136.1| similar to ORF2 [Mus musculus domesticus]
HG1001502N0_160000_gene_predictio n2	gi|2144100|pir||I64837 Set beta isoform - rat
HG1000066N0_160000_gene_predictio n1	gi|26337951|dbj|BAC32661.1| unnamed protein product [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000078N0_1000_gene_prediction1	gi|26346587|dbj|BAC36942.1| unnamed protein product [Mus musculus]
HG1000117N0_160000_gene_predictio n1	gi|20875580|ref|XP_131162.1| sorting nexin 7 [Mus musculus]
HG1000157N0_160000_gene_predictio n1	gi|26344914|dbj|BAC36106.1| unnamed protein product [Mus musculus]
HG1000194N0_160000_gene_predictio n1	gi|21313022|ref|NP_083674.1| RIKEN cDNA 5730496E24 [Mus musculus]
HG1000501N0_160000_gene_predictio n1	gi|27370478|ref|NP_766552.1| hypothetical protein E130310N06 [Mus musculus]
HG1000656N0_10000_gene_prediction 1	gi|12855078|dbj|BAB30210.1| unnamed protein product [Mus musculus]
HG1000656N0_10000_gene_prediction 2	gi|12855078|dbj|BAB30210.1| unnamed protein product [Mus musculus]
HG1000750N0_160000_gene_predictio n1	gi|26336392|dbj|BAC31881.1| unnamed protein product [Mus musculus]
HG1001012N0_160000_gene_predictio n1	gi|21312504|ref|NP_081554.1| RIKEN cDNA 2810432D09 [Mus musculus]
HG1001237N0_10000_gene_prediction 1	gi|20882986|ref|XP_126218.1| similar to Hermansky-Pudlak syndrome protein variant [Rattus norvegicus] [Mus musculus]
HG1000228N0_40000_gene_prediction 1	gi|26342390|dbj|BAC34857.1| unnamed protein product [Mus musculus]
HG1000228N0_20000_gene_prediction 1	gi|13507676|ref|NP_109647.1| pumilio 1 (Drosophila) [Mus musculus]
HG1000228N0_160000_gene_predictio n1	gi|13507676|ref|NP_109647.1| pumilio 1 (Drosophila) [Mus musculus]
HG1000390N0_160000_gene_predictio n1	gi|20892585|ref|XP_147977.1| RIKEN cDNA 2610001E17 [Mus musculus]
HG1000409N0_10000_gene_prediction 1	gi|26006245|dbj|BAC41465.1| mKIAA1047 protein [Mus musculus]
HG1000611N0_160000_gene_predictio n1	gi|6650539|gb|AAF21895.1|AF103877_1 epsilon-sarcoglycan [Mus musculus]
HG1000847N0_10000_gene_prediction 1	
HG1000015N0_0_gene_prediction1	gi|20467423|ref|NP_620570.1| chondroitin sulfate proteoglycan 4 [Mus musculus]
HG1000088N0_5000_gene_prediction1	gi|16741633|gb|AAH16619.1| pyruvate kinase 3 [Mus musculus]
HG1000143N0_10000_gene_prediction 1	gi|20896345|ref|XP_128324.1| carbonyl reductase 3 [Mus musculus]
HG1000167N0_5000_gene_prediction1	gi|12848663|dbj|BAB28043.1| unnamed protein product [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000243N0_5000_gene_prediction1	gi|8393534|ref|NP_058653.1| high mobility group protein 17 [Mus musculus]
HG1000825N0_160000_gene_prediction1	gi|21311983|ref|NP_080956.1| RIKEN cDNA 0610012C01 [Mus musculus]
HG1001019N0_1000_gene_prediction1	gi|26343769|dbj|BAC35541.1| unnamed protein product [Mus musculus]
HG1000044N0_160000_gene_prediction1	gi|15079309|gb|AAH11494.1| Similar to Myosin of the dilute-myosin-V family [Mus musculus]
HG1000100N0_10000_gene_prediction1	gi|4506127|ref|NP_002755.1| phosphoribosyl pyrophosphate synthetase 1 [Homo sapiens]
HG1000149N0_160000_gene_prediction1	gi|12834813|dbj|BAB23054.1| unnamed protein product [Mus musculus]
HG1000183N0_1000_gene_prediction1	gi|27370150|ref|NP_766364.1| hypothetical protein D630002G06 [Mus musculus]
HG1000183N0_160000_gene_prediction2	gi|27370150|ref|NP_766364.1| hypothetical protein D630002G06 [Mus musculus]
HG1000213N0_5000_gene_prediction1	gi|6753178|ref|NP_035923.1| breakpoint cluster region protein 1; barrier to autointegration factor [Mus musculus]
HG1000294N0_5000_gene_prediction1	gi|18390327|ref|NP_083908.1| protein phosphatase 1, regulatory (inhibitor) subunit 11; t-complex testis-expressed 5 [Mus musculus]
HG1000331N0_160000_gene_prediction1	gi|20840824|ref|XP_141031.1| similar to slit homolog 1 (Drosophila); slit (Drosophila) homolog 1; slit1 [Homo sapiens] [Mus musculus]
HG1000391N0_160000_gene_prediction2	gi|20887543|ref|XP_134475.1| RIKEN cDNA 2310022B05 [Mus musculus]
HG1000430N0_160000_gene_prediction1	gi|26382861|dbj|BAC25510.1| unnamed protein product [Mus musculus]
HG1000597N0_160000_gene_prediction1	gi|26325886|dbj|BAC26697.1| unnamed protein product [Mus musculus]
HG1000078N0_5000_gene_prediction1	gi|26346587|dbj|BAC36942.1| unnamed protein product [Mus musculus]
HG1000139N0_5000_gene_prediction1	gi|23597632|ref|XP_127052.2| similar to hypothetical protein FLJ13920 [Homo sapiens] [Mus musculus]
HG1000143N0_160000_gene_prediction1	gi|20896345|ref|XP_128324.1| carbonyl reductase 3 [Mus musculus]
HG1000162N0_160000_gene_prediction1	gi|20835770|ref|XP_132127.1| similar to 60S RIBOSOMAL PROTEIN L13 [Mus musculus]
HG1000168N0_160000_gene_predictio	gi|12841593|dbj|BAB25272.1| unnamed protein

FP ID	Fantom Top Hit Annotation
n1	product [Mus musculus]
HG1000187N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000247N0_160000_gene_predictio n1	gi|7656920|ref|NP_056547.1| axin2 [Mus musculus]
HG1000273N0_160000_gene_predictio n2	gi|25030042|ref|XP_207307.1| similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1000415N0_10000_gene_prediction 1	gi|9367840|emb|CAB97523.1| hypothetical protein, weakly similar to (AF102871) neuronal apoptosis inhibitory protein 2 [Mus musculus] [Homo sapiens]
HG1000539N0_160000_gene_predictio n1	gi|7521942|pir||T29096 gag polyprotein - murine endogenous retrovirus ERV-L
HG1000539N0_160000_gene_predictio n2	gi|7521942|pir||T29096 gag polyprotein - murine endogenous retrovirus ERV-L
HG1000560N0_160000_gene_predictio n1	gi|12860683|dbj|BAB32021.1| unnamed protein product [Mus musculus]
HG1000618N0_10000_gene_prediction 1	gi|26350749|dbj|BAC39011.1| unnamed protein product [Mus musculus]
HG1000740N0_160000_gene_predictio n1	gi|23601536|ref|XP_130965.2| Nice-4 protein homolog [Mus musculus]
HG1001197N0_160000_gene_predictio n1	gi|26327779|dbj|BAC27630.1| unnamed protein product [Mus musculus]
HG1000599N0_5000_gene_prediction1	gi|12836542|dbj|BAB23701.1| unnamed protein product [Mus musculus]
HG1000020N0_5000_gene_prediction1	gi|20887101|ref|XP_129228.1| similar to phosphoglucomutase 5 [Homo sapiens] [Mus musculus]
HG1000084N0_5000_gene_prediction1	gi|6678794|ref|NP_032953.1| mitogen activated protein kinase kinase 1; MAP kinase kinase 1; protein kinase, mitogen activated, kinase 1, p45 [Mus musculus]
HG1000135N0_5000_gene_prediction1	gi|21312189|ref|NP_081197.1| RIKEN cDNA 1810010A06 [Mus musculus]
HG1000169N0_20000_gene_prediction 1	gi|20886743|ref|XP_129211.1| phosphoserine aminotransferase [Mus musculus]
HG1000169N0_160000_gene_predictio n1	gi|20886743|ref|XP_129211.1| phosphoserine aminotransferase [Mus musculus]
HG1000189N0_160000_gene_predictio n1	gi|20879992|ref|XP_140210.1| similar to BG:DS01759.1 gene product [Drosophila melanogaster] [Mus musculus]
HG1000189N0_160000_gene_predictio n2	gi|20879992|ref|XP_140210.1| similar to BG:DS01759.1 gene product [Drosophila

FP ID	Fantom Top Hit Annotation
	melanogaster] [Mus musculus]
HG1000246N0_5000_gene_prediction1	gi|21450297|ref|NP_659157.1| UDP-GalNAc:polypeptide N-acetylgalactosaminyltransferase [Mus musculus]
HG1000248N0_0_gene_prediction1	gi|9790219|ref|NP_062745.1| destrin; Sid23p [Mus musculus]
HG1000288N0_10000_gene_prediction1	gi|20909512|ref|XP_153447.1| hypothetical protein XP_153447 [Mus musculus]
HG1000424N0_5000_gene_prediction1	gi|25031822|ref|XP_207741.1| hypothetical protein XP_207741 [Mus musculus]
HG1000443N0_40000_gene_prediction1	gi|26354072|dbj|BAC40666.1| unnamed protein product [Mus musculus]
HG1000590N0_1000_gene_prediction1	gi|26378096|dbj|BAB28595.2| unnamed protein product [Mus musculus]
HG1000626N0_160000_gene_prediction1	gi|9938030|ref|NP_064667.1| hypothetical protein, MNCb-4193; hypothetical protein MNCb-4193 [Mus musculus]
HG1000871N0_160000_gene_prediction1	gi|6752958|ref|NP_033742.1| activin A receptor, type II-like 1; activin receptor-like kinase-1 [Mus musculus]
HG1000959N0_10000_gene_prediction1	gi|22507385|ref|NP_081019.1| RIKEN cDNA 1110014F12 [Mus musculus]
HG1000961N0_160000_gene_prediction3	gi|20822904|ref|XP_131914.1| RIKEN cDNA 3110004O18 [Mus musculus]
HG1000974N0_5000_gene_prediction1	gi|26378096|dbj|BAB28595.2| unnamed protein product [Mus musculus]
HG1001045N0_160000_gene_prediction1	gi|25020138|ref|XP_207789.1| similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1001110N0_0_gene_prediction1	gi|23956080|ref|NP_058675.1| putative serine/threonine kinase [Mus musculus]
HG1001223N0_1000_gene_prediction1	gi|26339658|dbj|BAC33500.1| unnamed protein product [Mus musculus]
HG1001281N0_160000_gene_prediction1	gi|15431279|ref|NP_203538.1| dedicator of cyto-kinesis 2 [Mus musculus]
HG1001317N0_5000_gene_prediction1	gi|26327365|dbj|BAC27426.1| unnamed protein product [Mus musculus]
HG1001485N0_5000_gene_prediction1	gi|26327365|dbj|BAC27426.1| unnamed protein product [Mus musculus]
HG1000674N0_160000_gene_prediction1	gi|24211881|sp|Q8VCR8|KML2_MOUSE Myosin light chain kinase 2, skeletal/cardiac muscle (MLCK2)
HG1001017N0_10000_gene_prediction	gi|25019831|ref|XP_207463.1| similar to

FP ID	Fantom Top Hit Annotation
1	CD59B [Mus musculus]
HG1001017N0_1000_gene_prediction1	gi|25019831|ref|XP_207463.1| similar to CD59B [Mus musculus]
HG1000014N0_160000_gene_prediction2	gi|6680744|ref|NP_031528.1| ATPase, Na+/K+ transporting, beta 3 polypeptide; ATPase, Na+/K+ beta 3 polypeptide [Mus musculus]
HG1000043N0_160000_gene_prediction3	gi|26337385|dbj|BAC32378.1| unnamed protein product [Mus musculus]
HG1000052N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000084N0_5000_gene_prediction2	gi|6678794|ref|NP_032953.1| mitogen activated protein kinase kinase 1; MAP kinase kinase 1; protein kinase, mitogen activated, kinase 1, p45 [Mus musculus]
HG1000093N0_1000_gene_prediction1	gi|26350865|dbj|BAC39069.1| unnamed protein product [Mus musculus]
HG1000105N0_160000_gene_prediction1	gi|14198371|gb|AAH08247.1| Similar to cyclin B2 [Mus musculus]
HG1000157N0_1000_gene_prediction1	gi|5803225|ref|NP_006752.1| tyrosine 3/tryptophan 5 -monooxygenase activation protein, epsilon polypeptide; 14-3-3 epsilon; mitochondrial import stimulation factor L subunit; protein kinase C inhibitor protein-1 [Homo sapiens]
HG1000210N0_40000_gene_prediction1	gi|17160840|gb|AAH17597.1| RIKEN cDNA 5830401B18 gene [Mus musculus]
HG1000242N0_5000_gene_prediction1	gi|9789937|ref|NP_062768.1| DnaJ (Hsp40) homolog, subfamily A, member 2; DNA J protein [Mus musculus]
HG1000243N0_5000_gene_prediction2	gi|8393534|ref|NP_058653.1| high mobility group protein 17 [Mus musculus]
HG1000256N0_160000_gene_prediction1	gi|13959400|sp|Q9R0Y5|KAD1_MOUSE Adenylate kinase isoenzyme 1 (ATP-AMP transphosphorylase) (AK1) (Myokinase)
HG1000279N0_0_gene_prediction1	gi|15617203|ref|NP_254279.1| chloride intracellular channel 1 [Mus musculus]
HG1000280N0_5000_gene_prediction1	gi|7106337|ref|NP_034796.1| keratin complex-1, gene C29 [Mus musculus]
HG1000280N0_5000_gene_prediction2	gi|7106337|ref|NP_034796.1| keratin complex-1, gene C29 [Mus musculus]
HG1000282N0_160000_gene_prediction1	gi|20902823|ref|XP_128021.1| similar to Mitochondrial import receptor subunit TOM22 homolog (Translocase of outer membrane 22 kDa subunit homolog) (hTom22) (1C9-2) [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000292N0_160000_gene_predictio n1	gi|6981488|ref|NP_037356.1| ribosomal protein S26 [Rattus norvegicus]
HG1000313N0_160000_gene_predictio n1	gi|4506283|ref|NP_003454.1| protein tyrosine phosphatase type IVA, member 1; Protein tyrosine phosphatase IVA1 [Homo sapiens]
HG1000330N0_20000_gene_prediction 1	gi|22122511|ref|NP_666146.1| hypothetical protein MGC30562 [Mus musculus]
HG1000339N0_160000_gene_predictio n1	gi|26350551|dbj|BAC38915.1| unnamed protein product [Mus musculus]
HG1000340N0_160000_gene_predictio n1	gi|20912842|ref|XP_126689.1| RIKEN cDNA 3300001P08 [Mus musculus]
HG1000344N0_160000_gene_predictio n1	gi|21450239|ref|NP_659092.1| hypothetical protein MGC27983 [Mus musculus]
HG1000365N0_20000_gene_prediction 1	gi|25046794|ref|XP_207489.1| similar to RNP particle component [Mus musculus]
HG1000384N0_160000_gene_predictio n1	gi|20909520|ref|XP_126941.1| RIKEN cDNA 2600011C06 [Mus musculus]
HG1000448N0_160000_gene_predictio n1	gi|6678247|ref|NP_033358.1| transcription factor 7-like 1 [Mus musculus]
HG1000482N0_160000_gene_predictio n1	gi|26334795|dbj|BAC31098.1| unnamed protein product [Mus musculus]
HG1000486N0_20000_gene_prediction 1	gi|26350551|dbj|BAC38915.1| unnamed protein product [Mus musculus]
HG1000506N0_160000_gene_predictio n1	gi|20909520|ref|XP_126941.1| RIKEN cDNA 2600011C06 [Mus musculus]
HG1000518N0_160000_gene_predictio n1	gi|26351279|dbj|BAC39276.1| unnamed protein product [Mus musculus]
HG1000550N0_160000_gene_predictio n1	gi|20909520|ref|XP_126941.1| RIKEN cDNA 2600011C06 [Mus musculus]
HG1000556N0_160000_gene_predictio n1	gi|25031497|ref|XP_207552.1| similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1000588N0_160000_gene_predictio n1	gi|13277747|gb|AAH03768.1| interferon-induced protein with tetratricopeptide repeats 1 [Mus musculus]
HG1000600N0_160000_gene_predictio n1	gi|20863376|ref|XP_134148.1| similar to hypothetical protein [Macaca fascicularis] [Mus musculus]
HG1000647N0_160000_gene_predictio n1	gi|9506517|ref|NP_062338.1| cytotoxic and regulatory T cell molecule; class I-restricted T cell-associated molecule [Mus musculus]
HG1000648N0_160000_gene_predictio n1	gi|20900199|ref|XP_128639.1| RIKEN cDNA 2810055C19 [Mus musculus]
HG1000688N0_160000_gene_predictio	gi|26327707|dbj|BAC27597.1| unnamed protein

FP ID	Fantom Top Hit Annotation
n1	product [Mus musculus]
HG1000696N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000788N0_160000_gene_predictio n1	gi|20847912|ref|XP_144610.1| similar to KIAA1904 protein [Homo sapiens] [Mus musculus]
HG1000874N0_160000_gene_predictio n1	gi|20342176|ref|XP_110490.1| similar to hypothetical protein MGC955 [Homo sapiens] [Mus musculus]
HG1000902N0_20000_gene_prediction 1	gi|6753324|ref|NP_033968.1| chaperonin subunit 6a (zeta); chaperonin containing TCP-1 [Mus musculus]
HG1000902N0_160000_gene_predictio n2	gi|6753324|ref|NP_033968.1| chaperonin subunit 6a (zeta); chaperonin containing TCP-1 [Mus musculus]
HG1000902N0_1000_gene_prediction1	gi|6753324|ref|NP_033968.1| chaperonin subunit 6a (zeta); chaperonin containing TCP-1 [Mus musculus]
HG1000904N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000966N0_1000_gene_prediction1	gi|22122617|ref|NP_666215.1| hypothetical protein MGC25511 [Mus musculus]
HG1000966N0_5000_gene_prediction1	gi|22122617|ref|NP_666215.1| hypothetical protein MGC25511 [Mus musculus]
HG1000994N0_160000_gene_predictio n1	gi|12855175|dbj|BAB30238.1| unnamed protein product [Mus musculus]
HG1001014N0_160000_gene_predictio n3	gi|26337385|dbj|BAC32378.1| unnamed protein product [Mus musculus]
HG1001041N0_5000_gene_prediction1	gi|25071304|ref|XP_146497.3| similar to protein serine kinase Pskh1 [Mus musculus]
HG1001337N0_160000_gene_predictio n1	gi|27369704|ref|NP_766096.1| hypothetical protein 6030499O08 [Mus musculus]
HG1001417N0_5000_gene_prediction1	gi|26349767|dbj|BAC38523.1| unnamed protein product [Mus musculus]
HG1001485N0_160000_gene_predictio n1	gi|7513636|pir||T30805 dutt1 protein - mouse
HG1000151N0_160000_gene_predictio n1	gi|18044328|gb|AAH19573.1| Unknown (protein for IMAGE:3990036) [Mus musculus]
HG1000330N0_160000_gene_predictio n3	gi|25029811|ref|XP_207217.1| similar to ORF2 [Mus musculus domesticus]
HG1000957N0_20000_gene_prediction 1	gi|25024769|ref|XP_207136.1| similar to ORF2 [Mus musculus domesticus]
HG1000960N0_0_gene_prediction1	gi|20908689|ref|XP_127449.1| RIKEN cDNA 4632401C08 [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000960N0_0_gene_prediction2	gi|20908689|ref|XP_127449.1| RIKEN cDNA 4632401C08 [Mus musculus]
HG1001280N0_20000_gene_prediction1	gi|26336763|dbj|BAC32064.1| unnamed protein product [Mus musculus]
HG1001502N0_160000_gene_prediction1	gi|27370240|ref|NP_766415.1| hypothetical protein 4732490P18 [Mus musculus]
HG1000003N0_10000_gene_prediction1	gi|13624305|ref|NP_112440.1| procollagen, type II, alpha 1 [Mus musculus]
HG1000041N0_160000_gene_prediction1	gi|26390169|dbj|BAC25854.1| unnamed protein product [Mus musculus]
HG1000043N0_160000_gene_prediction2	gi|26337385|dbj|BAC32378.1| unnamed protein product [Mus musculus]
HG1000044N0_5000_gene_prediction1	gi|15079309|gb|AAH11494.1| Similar to Myosin of the dilute-myosin-V family [Mus musculus]
HG1000051N0_160000_gene_prediction1	gi|14250190|gb|AAH08515.1| interferon regulatory factor 6 [Mus musculus]
HG1000057N0_160000_gene_prediction1	gi|6755040|ref|NP_035202.1| profilin 1; actin binding protein [Mus musculus]
HG1000060N0_160000_gene_prediction1	gi|6755901|ref|NP_035783.1| tubulin, alpha 1; tubulin alpha 1 [Mus musculus]
HG1000061N0_10000_gene_prediction1	gi|20827552|ref|XP_130234.1| expressed sequence AW610751 [Mus musculus]
HG1000079N0_160000_gene_prediction1	gi|20887309|ref|XP_129200.1| adenylate kinase 3 alpha like [Mus musculus]
HG1000098N0_160000_gene_prediction1	gi|26340666|dbj|BAC33995.1| unnamed protein product [Mus musculus]
HG1000105N0_5000_gene_prediction1	gi|12850600|dbj|BAB28785.1| unnamed protein product [Mus musculus]
HG1000121N0_160000_gene_prediction1	gi|26346402|dbj|BAC36852.1| unnamed protein product [Mus musculus]
HG1000131N0_160000_gene_prediction1	gi|26329183|dbj|BAC28330.1| unnamed protein product [Mus musculus]
HG1000134N0_160000_gene_prediction1	gi|12860377|dbj|BAB31934.1| unnamed protein product [Mus musculus]
HG1000134N0_160000_gene_prediction2	gi|12860377|dbj|BAB31934.1| unnamed protein product [Mus musculus]
HG1000136N0_160000_gene_prediction1	gi|26389519|dbj|BAC25745.1| unnamed protein product [Mus musculus]
HG1000147N0_160000_gene_prediction1	gi|3717978|emb|CAA73041.1| 5S ribosomal protein [Mus musculus]
HG1000166N0_160000_gene_prediction1	gi|20908717|ref|XP_127445.1| similar to flavoprotein subunit of succinate-ubiquinone reductase [Rattus norvegicus] [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000172N0_1000_gene_prediction1	gi|6681095|ref|NP_031834.1| cytochrome c, somatic [Mus musculus]
HG1000172N0_1000_gene_prediction2	gi|6681095|ref|NP_031834.1| cytochrome c, somatic [Mus musculus]
HG1000175N0_5000_gene_prediction1	gi|26354216|dbj|BAC40736.1| unnamed protein product [Mus musculus]
HG1000175N0_10000_gene_prediction 1	gi|26354216|dbj|BAC40736.1| unnamed protein product [Mus musculus]
HG1000175N0_160000_gene_predictio n1	gi|26354216|dbj|BAC40736.1| unnamed protein product [Mus musculus]
HG1000175N0_1000_gene_prediction1	gi|26354216|dbj|BAC40736.1| unnamed protein product [Mus musculus]
HG1000192N0_160000_gene_predictio n1	gi|10946614|ref|NP_067287.1| WD repeat domain 12; nuclear protein Ytm1 [Mus musculus]
HG1000193N0_160000_gene_predictio n2	gi|21728370|ref|NP_080178.1| RIKEN cDNA 1500009M05 [Mus musculus]
HG1000195N0_160000_gene_predictio n1	gi|17390530|gb|AAH18231.1| Unknown (protein for MGC:19236) [Mus musculus]
HG1000197N0_160000_gene_predictio n1	gi|21450185|ref|NP_659063.1| hypothetical protein MGC28186 [Mus musculus]
HG1000202N0_20000_gene_prediction 1	gi|26331946|dbj|BAC29703.1| unnamed protein product [Mus musculus]
HG1000210N0_20000_gene_prediction 1	gi|17160840|gb|AAH17597.1| RIKEN cDNA 5830401B18 gene [Mus musculus]
HG1000218N0_1000_gene_prediction1	gi|6681015|ref|NP_031789.1| cysteine rich intestinal protein [Mus musculus]
HG1000218N0_160000_gene_predictio n1	gi|6681015|ref|NP_031789.1| cysteine rich intestinal protein [Mus musculus]
HG1000218N0_10000_gene_prediction 1	gi|6681015|ref|NP_031789.1| cysteine rich intestinal protein [Mus musculus]
HG1000222N0_1000_gene_prediction1	gi|13385054|ref|NP_079873.1| RIKEN cDNA 2700033I16 [Mus musculus]
HG1000233N0_1000_gene_prediction1	gi|12847362|dbj|BAB27541.1| unnamed protein product [Mus musculus]
HG1000234N0_1000_gene_prediction1	gi|12847362|dbj|BAB27541.1| unnamed protein product [Mus musculus]
HG1000234N0_160000_gene_predictio n1	gi|12847362|dbj|BAB27541.1| unnamed protein product [Mus musculus]
HG1000238N0_160000_gene_predictio n2	gi|6671549|ref|NP_031479.1| anti-oxidant protein 2; acidic calcium-independent phospholipase A2; peroxiredoxin 5; 1-Cys Prx [Mus musculus]
HG1000240N0_160000_gene_predictio	gi|26328673|dbj|BAC28075.1| unnamed protein

FP ID	Fantom Top Hit Annotation
n1	product [Mus musculus]
HG1000245N0_160000_gene_predictio n1	gi|12850132|dbj|BAB28604.1| unnamed protein product [Mus musculus]
HG1000245N0_5000_gene_prediction1	gi|12850132|dbj|BAB28604.1| unnamed protein product [Mus musculus]
HG1000249N0_10000_gene_prediction 1	gi|6754654|ref|NP_034905.1| mannose binding lectin, liver (A) [Mus musculus]
HG1000251N0_160000_gene_predictio n1	gi|20881913|ref|XP_126211.1| Dullard homolog [Mus musculus]
HG1000252N0_5000_gene_prediction1	gi|20825536|ref|XP_129507.1| ring finger protein 2 [Mus musculus]
HG1000254N0_160000_gene_predictio n1	gi|13385058|ref|NP_079878.1| hypothetical protein D10Ertd718e [Mus musculus]
HG1000262N0_160000_gene_predictio n1	gi|21312163|ref|NP_082683.1| RIKEN cDNA 2900054P12 [Mus musculus]
HG1000264N0_1000_gene_prediction1	gi|21624617|ref|NP_081018.1| RIKEN cDNA 1110007M04 [Mus musculus]
HG1000264N0_1000_gene_prediction2	gi|21624617|ref|NP_081018.1| RIKEN cDNA 1110007M04 [Mus musculus]
HG1000270N0_20000_gene_prediction 1	gi|12844196|dbj|BAB26273.1| unnamed protein product [Mus musculus]
HG1000270N0_1000_gene_prediction1	gi|12852884|dbj|BAB29566.1| unnamed protein product [Mus musculus]
HG1000274N0_160000_gene_predictio n1	gi|26347831|dbj|BAC37564.1| unnamed protein product [Mus musculus]
HG1000276N0_160000_gene_predictio n1	gi|19527228|ref|NP_598768.1| DNA segment, Chr 10, ERATO Doi 214, expressed [Mus musculus]
HG1000276N0_5000_gene_prediction1	gi|19527228|ref|NP_598768.1| DNA segment, Chr 10, ERATO Doi 214, expressed [Mus musculus]
HG1000278N0_5000_gene_prediction1	gi|19527026|ref|NP_598568.1| expressed sequence AA959742 [Mus musculus]
HG1000280N0_160000_gene_predictio n1	gi|7106337|ref|NP_034796.1| keratin complex-1, gene C29 [Mus musculus]
HG1000280N0_1000_gene_prediction1	gi|7106337|ref|NP_034796.1| keratin complex-1, gene C29 [Mus musculus]
HG1000280N0_160000_gene_predictio n2	gi|7106337|ref|NP_034796.1| keratin complex-1, gene C29 [Mus musculus]
HG1000280N0_1000_gene_prediction2	gi|7106337|ref|NP_034796.1| keratin complex-1, gene C29 [Mus musculus]
HG1000305N0_5000_gene_prediction1	gi|27369902|ref|NP_766218.1| hypothetical protein A530095G11 [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000305N0_5000_gene_prediction2	gi|27369902|ref|NP_766218.1| hypothetical protein A530095G11 [Mus musculus]
HG1000307N0_160000_gene_predictio n1	gi|8393853|ref|NP_058614.1| nudix (nucleoside diphosphate linked moiety X)-type motif 5 [Mus musculus]
HG1000334N0_160000_gene_predictio n1	gi|20888553|ref|XP_134832.1| similar to Probable serine/threonine protein kinase SNF1LK [Mus musculus]
HG1000335N0_160000_gene_predictio n1	gi|20888553|ref|XP_134832.1| similar to Probable serine/threonine protein kinase SNF1LK [Mus musculus]
HG1000337N0_5000_gene_prediction1	gi|12851918|dbj|BAB29207.1| unnamed protein product [Mus musculus]
HG1000343N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000343N0_160000_gene_predictio n2	gi|13386340|ref|NP_083008.1| RIKEN cDNA 4632428N05 [Mus musculus]
HG1000369N0_160000_gene_predictio n1	gi|12837873|dbj|BAB23982.1| unnamed protein product [Mus musculus]
HG1000372N0_160000_gene_predictio n1	gi|20913947|ref|XP_126555.1| RIKEN cDNA 1190006K01 [Mus musculus]
HG1000378N0_160000_gene_predictio n1	gi|26348995|dbj|BAC38137.1| unnamed protein product [Mus musculus]
HG1000387N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000387N0_160000_gene_predictio n2	gi|26382861|dbj|BAC25510.1| unnamed protein product [Mus musculus]
HG1000397N0_5000_gene_prediction1	gi|20836469|ref|XP_129717.1| hypothetical protein XP_129717 [Mus musculus]
HG1000408N0_160000_gene_predictio n1	
HG1000414N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000431N0_160000_gene_predictio n1	gi|8394057|ref|NP_058565.1| low density lipoprotein receptor-related protein 4; low density lipoprotein-related protein 4; Low Density Lipoprotein Receptor Related Protein 4; corin [Mus musculus]
HG1000439N0_160000_gene_predictio n1	gi|12851918|dbj|BAB29207.1| unnamed protein product [Mus musculus]
HG1000449N0_20000_gene_prediction 1	gi|25025117|ref|XP_207206.1| similar to transcription factor-like nuclear regulator; putative transcription regulation nuclear protein; putative transcription factor-like

FP ID	Fantom Top Hit Annotation
	nuclear regulator; TATA box binding protein (TBP)-associated factor, RNA polymerase III, GTF3B subunit 1; ... [Mus musculus]
HG1000457N0_160000_gene_predictio n1	gi|20824761|ref|XP_133346.1| liver-specific bHLH-Zip transcription factor [Mus musculus]
HG1000458N0_160000_gene_predictio n1	gi|12841242|dbj|BAB25129.1| unnamed protein product [Mus musculus]
HG1000461N0_160000_gene_predictio n1	gi|25032310|ref|XP_205729.1| hypothetical protein XP_205729 [Mus musculus]
HG1000463N0_160000_gene_predictio n1	gi|12861068|dbj|BAB32114.1| unnamed protein product [Mus musculus]
HG1000463N0_160000_gene_predictio n2	gi|13249351|ref|NP_076402.1| inositol-requiring 1 alpha (yeast) [Mus musculus]
HG1000476N0_160000_gene_predictio n1	gi|26332657|dbj|BAC30046.1| unnamed protein product [Mus musculus]
HG1000481N0_160000_gene_predictio n1	gi|21311873|ref|NP_077181.1| RIKEN cDNA 0610007A03 [Mus musculus]
HG1000530N0_160000_gene_predictio n1	gi|20860491|ref|XP_153755.1| hypothetical protein XP_153755 [Mus musculus]
HG1000556N0_160000_gene_predictio n2	gi|25031497|ref|XP_207552.1| similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1000584N0_160000_gene_predictio n1	gi|27370500|ref|NP_766581.1| hypothetical protein D230008H22 [Mus musculus]
HG1000587N0_160000_gene_predictio n1	gi|23682449|ref|XP_158842.2| hypothetical protein XP_158842 [Mus musculus]
HG1000592N0_160000_gene_predictio n1	gi|26349599|dbj|BAC38439.1| unnamed protein product [Mus musculus]
HG1000594N0_160000_gene_predictio n1	gi|22095015|ref|NP_084065.1| RIKEN cDNA 0610013I17 [Mus musculus]
HG1000594N0_160000_gene_predictio n2	gi|22095015|ref|NP_084065.1| RIKEN cDNA 0610013I17 [Mus musculus]
HG1000608N0_160000_gene_predictio n1	gi|20345223|ref|XP_109778.1| similar to Neurabin-II (Neural tissue-specific F-actin binding protein II) (Protein phosphatase 1 regulatory subunit 9B) (Spinophilin) (p130) (PP1bp134) [Mus musculus]
HG1000615N0_160000_gene_predictio n1	gi|7710032|ref|NP_057928.1| growth factor receptor bound protein 14 [Mus musculus]
HG1000620N0_160000_gene_predictio n1	gi|25052462|ref|XP_138105.3| similar to TAR DNA-binding protein-43 (TDP-43) [Mus musculus]
HG1000621N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]

FP ID	Fantom Top Hit Annotation
HG1000621N0_160000_gene_prediction3	gi|26382861|dbj|BAC25510.1| unnamed protein product [Mus musculus]
HG1000631N0_40000_gene_prediction1	gi|6681283|ref|NP_031938.1| epidermal growth factor receptor; avian erythroblastic leukemia viral (v-erb-b) oncogene homolog [Mus musculus]
HG1000652N0_160000_gene_prediction1	gi|25030122|ref|XP_207332.1| similar to endonuclease/reverse transcriptase [Mus musculus]
HG1000663N0_160000_gene_prediction1	gi|20915416|ref|XP_162987.1| hypothetical protein XP_162987 [Mus musculus]
HG1000686N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000700N0_160000_gene_prediction1	gi|16508047|gb|AAL17972.1| pORF2 [Mus musculus domesticus]
HG1000701N0_160000_gene_prediction1	gi|26327167|dbj|BAC27327.1| unnamed protein product [Mus musculus]
HG1000709N0_160000_gene_prediction1	gi|220579|dbj|BAA00448.1| open reading frame (196 AA) [Mus musculus]
HG1000712N0_160000_gene_prediction1	gi|12841826|dbj|BAB25366.1| unnamed protein product [Mus musculus]
HG1000720N0_160000_gene_prediction1	gi|7657415|ref|NP_035986.2| odd Oz/ten-m homolog 2 (Drosophila); odd Oz/ten-m homolog 3 (Drosophila) [Mus musculus]
HG1000727N0_160000_gene_prediction1	gi|26335645|dbj|BAC31523.1| unnamed protein product [Mus musculus]
HG1000743N0_160000_gene_prediction2	gi|26338834|dbj|BAC33088.1| unnamed protein product [Mus musculus]
HG1000767N0_5000_gene_prediction1	gi|12851918|dbj|BAB29207.1| unnamed protein product [Mus musculus]
HG1000786N0_160000_gene_prediction2	gi|6678303|ref|NP_033386.1| transcription factor A, mitochondrial [Mus musculus]
HG1000822N0_160000_gene_prediction1	gi|6680195|ref|NP_032255.1| histone deacetylase 2; DNA segment, Chr 10, Wayne State University 179, expressed [Mus musculus]
HG1000829N0_160000_gene_prediction1	gi|21450159|ref|NP_659049.1| cDNA sequence BC024131; hypothetical protein MGC37896 [Mus musculus]
HG1000848N0_160000_gene_prediction1	gi|26350995|dbj|BAC39134.1| unnamed protein product [Mus musculus]
HG1000860N0_160000_gene_prediction1	gi|26325678|dbj|BAC26593.1| unnamed protein product [Mus musculus]
HG1000898N0_10000_gene_prediction	gi|21450209|ref|NP_659075.1| hypothetical

FP ID	Fantom Top Hit Annotation
1	protein MGC25509 [Mus musculus]
HG1000898N0_160000_gene_prediction1	gi|21450209|ref|NP_659075.1| hypothetical protein MGC25509 [Mus musculus]
HG1000898N0_20000_gene_prediction1	gi|21450209|ref|NP_659075.1| hypothetical protein MGC25509 [Mus musculus]
HG1000902N0_160000_gene_prediction1	gi|21450209|ref|NP_659075.1| hypothetical protein MGC25509 [Mus musculus]
HG1000904N0_160000_gene_prediction3	gi|6753324|ref|NP_033968.1| chaperonin subunit 6a (zeta); chaperonin containing TCP-1 [Mus musculus]
HG1000906N0_20000_gene_prediction1	gi|20344324|ref|XP_109683.1| RIKEN cDNA 1810027O10 [Mus musculus]
HG1000906N0_160000_gene_prediction1	gi|26346114|dbj|BAC36708.1| unnamed protein product [Mus musculus]
HG1000921N0_5000_gene_prediction1	gi|26346114|dbj|BAC36708.1| unnamed protein product [Mus musculus]
HG1000938N0_10000_gene_prediction1	gi|26350775|dbj|BAC39024.1| unnamed protein product [Mus musculus]
HG1000952N0_160000_gene_prediction1	gi|26339054|dbj|BAC33198.1| unnamed protein product [Mus musculus]
HG1000961N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000961N0_160000_gene_prediction2	gi|25051287|ref|XP_146665.3| similar to KIAA0877 protein [Homo sapiens] [Mus musculus]
HG1001000N0_160000_gene_prediction2	gi|20859143|ref|XP_127126.1| similar to eukaryotic initiation factor 5 [Rattus norvegicus] [Mus musculus]
HG1001003N0_160000_gene_prediction1	gi|19527072|ref|NP_598613.1| expressed sequence AW555139 [Mus musculus]
HG1001007N0_160000_gene_prediction1	gi|13277825|gb|AAH03796.1| Similar to lymphocyte specific 1 [Mus musculus]
HG1001009N0_0_gene_prediction1	gi|26334641|dbj|BAC31021.1| unnamed protein product [Mus musculus]
HG1001014N0_160000_gene_prediction2	gi|26329567|dbj|BAC28522.1| unnamed protein product [Mus musculus]
HG1001017N0_40000_gene_prediction1	gi|26337385|dbj|BAC32378.1| unnamed protein product [Mus musculus]
HG1001017N0_20000_gene_prediction1	gi|25019831|ref|XP_207463.1| similar to CD59B [Mus musculus]
HG1001144N0_160000_gene_prediction1	gi|25019831|ref|XP_207463.1| similar to CD59B [Mus musculus]
HG1001172N0_160000_gene_prediction2	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]

FP ID	Fantom Top Hit Annotation
HG1001214N0_20000_gene_prediction 1	gi|26340706|dbj|BAC34015.1| unnamed protein product [Mus musculus]
HG1001229N0_160000_gene_predictio n1	
HG1001253N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1001253N0_160000_gene_predictio n2	gi|26326251|dbj|BAC26869.1| unnamed protein product [Mus musculus]
HG1001267N0_160000_gene_predictio n1	gi|26326251|dbj|BAC26869.1| unnamed protein product [Mus musculus]
HG1001289N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1001343N0_10000_gene_prediction 1	gi|26333317|dbj|BAC30376.1| unnamed protein product [Mus musculus]
HG1001343N0_160000_gene_predictio n1	gi|6755060|ref|NP_035214.1| phosphatidylinositol 3-kinase, C2 domain containing, gamma polypeptide [Mus musculus]
HG1001390N0_160000_gene_predictio n1	gi|6755060|ref|NP_035214.1| phosphatidylinositol 3-kinase, C2 domain containing, gamma polypeptide [Mus musculus]
HG1001468N0_160000_gene_predictio n1	gi|6680083|ref|NP_032189.1| growth factor receptor bound protein 2 [Mus musculus]
HG1001508N0_160000_gene_predictio n2	gi|25030495|ref|XP_205178.1| similar to bA130N24.1 (novel protein similar to REV3L (REV3 (yeast homolog)-like, catalytic subunit of DNA polymerase zeta) (POLZ)) [Homo sapiens] [Mus musculus]
HG1000084N0_160000_gene_predictio n1	gi|26382861|dbj|BAC25510.1| unnamed protein product [Mus musculus]
HG1000084N0_160000_gene_predictio n2	gi|25031822|ref|XP_207741.1| hypothetical protein XP_207741 [Mus musculus]
HG1000209N0_160000_gene_predictio n1	gi|25031822|ref|XP_207741.1| hypothetical protein XP_207741 [Mus musculus]
HG1000382N0_160000_gene_predictio n1	gi|20858167|ref|XP_125585.1| similar to PTD013 protein; CGI-24 protein [Mus musculus]
HG1000591N0_160000_gene_predictio n1	gi|6678716|ref|NP_032539.1| low density lipoprotein receptor-related protein 5; low density lipoprotein-related protein 5 [Mus musculus]
HG1000904N0_160000_gene_predictio n4	gi|26330005|dbj|BAC28741.1| unnamed protein product [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000005N0_160000_gene_prediction1	gi|20835832|ref|XP_129684.1| complement receptor 2 [Mus musculus]
HG1000014N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000015N0_160000_gene_prediction1	gi|6680744|ref|NP_031528.1| ATPase, Na+/K+ transporting, beta 3 polypeptide; ATPase, Na+/K+ beta 3 polypeptide [Mus musculus]
HG1000015N0_20000_gene_prediction1	gi|20467423|ref|NP_620570.1| chondroitin sulfate proteoglycan 4 [Mus musculus]
HG1000015N0_5000_gene_prediction1	gi|20467423|ref|NP_620570.1| chondroitin sulfate proteoglycan 4 [Mus musculus]
HG1000015N0_160000_gene_prediction2	gi|20467423|ref|NP_620570.1| chondroitin sulfate proteoglycan 4 [Mus musculus]
HG1000020N0_160000_gene_prediction1	gi|20467423|ref|NP_620570.1| chondroitin sulfate proteoglycan 4 [Mus musculus]
HG1000020N0_5000_gene_prediction2	gi|26330706|dbj|BAC29083.1| unnamed protein product [Mus musculus]
HG1000024N0_10000_gene_prediction1	gi|20887101|ref|XP_129228.1| similar to phosphoglucomutase 5 [Homo sapiens] [Mus musculus]
HG1000026N0_160000_gene_prediction1	gi|12853786|dbj|BAB29848.1| unnamed protein product [Mus musculus]
HG1000030N0_160000_gene_prediction1	gi|9506367|ref|NP_062425.1| ATP-binding cassette, sub-family B, member 10; ATP-binding cassette, sub-family B (MDR/TAP), member 12; Abc-mitochondrial erythroid [Mus musculus]
HG1000039N0_160000_gene_prediction1	gi|26006203|dbj|BAC41444.1| mKIAA0696 protein [Mus musculus]
HG1000041N0_5000_gene_prediction1	gi|7106453|ref|NP_035897.1| zinc finger RNA binding protein [Mus musculus]
HG1000043N0_160000_gene_prediction1	gi|26390169|dbj|BAC25854.1| unnamed protein product [Mus musculus]
HG1000043N0_5000_gene_prediction1	gi|26337385|dbj|BAC32378.1| unnamed protein product [Mus musculus]
HG1000044N0_20000_gene_prediction1	gi|26337385|dbj|BAC32378.1| unnamed protein product [Mus musculus]
HG1000052N0_160000_gene_prediction2	gi|15079309|gb|AAH11494.1| Similar to Myosin of the dilute-myosin-V family [Mus musculus]
HG1000052N0_10000_gene_prediction1	gi|26324852|dbj|BAC26180.1| unnamed protein product [Mus musculus]
HG1000052N0_20000_gene_prediction1	gi|26324852|dbj|BAC26180.1| unnamed protein product [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000058N0_10000_gene_prediction 1	gi|26324852|dbj|BAC26180.1| unnamed protein product [Mus musculus]
HG1000061N0_5000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000065N0_5000_gene_prediction1	gi|5031571|ref|NP_005713.1| actin-related protein 2; ARP2 (actin-related protein 2, yeast) homolog [Homo sapiens]
HG1000065N0_10000_gene_prediction 1	gi|13386220|ref|NP_081610.1| RIKEN cDNA 2210414H16 [Mus musculus]
HG1000065N0_160000_gene_predictio n1	gi|13386220|ref|NP_081610.1| RIKEN cDNA 2210414H16 [Mus musculus]
HG1000068N0_160000_gene_predictio n1	gi|13386220|ref|NP_081610.1| RIKEN cDNA 2210414H16 [Mus musculus]
HG1000070N0_0_gene_prediction1	gi|26326191|dbj|BAC26839.1| unnamed protein product [Mus musculus]
HG1000073N0_20000_gene_prediction 1	gi|21595527|gb|AAH32275.1| Similar to receptor-like tyrosine kinase [Mus musculus]
HG1000075N0_160000_gene_predictio n1	gi|26326407|dbj|BAC26947.1| unnamed protein product [Mus musculus]
HG1000076N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000081N0_160000_gene_predictio n1	gi|4502549|ref|NP_001734.1| calmodulin 2 (phosphorylase kinase, delta); phosphorylase kinase delta [Homo sapiens]
HG1000106N0_160000_gene_predictio n1	gi|6680305|ref|NP_032328.1| heat shock protein, 84 kDa 1 [Mus musculus]
HG1000107N0_160000_gene_predictio n1	gi|6681225|ref|NP_031905.1| developmentally regulated GTP binding protein 1; developmentally regulated GTP-binding protein 1 [Mus musculus]
HG1000109N0_0_gene_prediction1	gi|6754774|ref|NP_034986.1| myosin heavy chain, cardiac muscle, adult; alpha cardiac MHC; alpha myosin [Mus musculus]
HG1000112N0_160000_gene_predictio n1	gi|23956080|ref|NP_058675.1| putative serine/threonine kinase [Mus musculus]
HG1000116N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000126N0_160000_gene_predictio n1	gi|6680305|ref|NP_032328.1| heat shock protein, 84 kDa 1 [Mus musculus]
HG1000130N0_160000_gene_predictio n1	gi|20825377|ref|XP_143696.1| similar to hypothetical protein dJ122O8.2 [Homo sapiens] [Mus musculus]
HG1000132N0_160000_gene_predictio n1	gi|6754208|ref|NP_034569.1| high mobility group box 1; high mobility group protein 1

FP ID	Fantom Top Hit Annotation
	[Mus musculus]
HG1000133N0_160000_gene_prediction1	gi|26347765|dbj|BAC37531.1| unnamed protein product [Mus musculus]
HG1000134N0_20000_gene_prediction1	gi|26382599|dbj|BAB22733.2| unnamed protein product [Mus musculus]
HG1000134N0_20000_gene_prediction2	gi|26353738|dbj|BAC40499.1| unnamed protein product [Mus musculus]
HG1000142N0_160000_gene_prediction1	gi|26353738|dbj|BAC40499.1| unnamed protein product [Mus musculus]
HG1000144N0_20000_gene_prediction1	gi|6679108|ref|NP_032748.1| nucleophosmin 1; nucleolar protein NO38 [Mus musculus]
HG1000145N0_160000_gene_prediction1	gi|6677779|ref|NP_033107.1| ribosomal protein L28; DNA segment, Chr 7, Wayne State University 21, expressed [Mus musculus]
HG1000146N0_160000_gene_prediction1	gi|6677779|ref|NP_033107.1| ribosomal protein L28; DNA segment, Chr 7, Wayne State University 21, expressed [Mus musculus]
HG1000150N0_10000_gene_prediction1	gi|3717978|emb|CAA73041.1| 5S ribosomal protein [Mus musculus]
HG1000152N0_160000_gene_prediction1	gi|11037798|ref|NP_067621.1| dynactin 5; dynactin 4; p25 dynactin subunit [Mus musculus]
HG1000161N0_160000_gene_prediction1	gi|21536242|ref|NP_573499.1| glucocorticoid induced transcript 1; testhymin; thymocyte/spermatocyte selection 1 [Mus musculus]
HG1000163N0_160000_gene_prediction1	gi|20819730|ref|XP_129359.1| hypothetical protein XP_129359 [Mus musculus]
HG1000164N0_5000_gene_prediction1	gi|20835770|ref|XP_132127.1| similar to 60S RIBOSOMAL PROTEIN L13 [Mus musculus]
HG1000165N0_1000_gene_prediction1	gi|26340448|dbj|BAC33887.1| unnamed protein product [Mus musculus]
HG1000166N0_160000_gene_prediction2	gi|26353666|dbj|BAC40463.1| unnamed protein product [Mus musculus]
HG1000167N0_160000_gene_prediction1	gi|27369878|ref|NP_766203.1| hypothetical protein 5330403K09 [Mus musculus]
HG1000171N0_40000_gene_prediction1	gi|26354683|dbj|BAC40968.1| unnamed protein product [Mus musculus]
HG1000171N0_160000_gene_prediction1	gi|26325838|dbj|BAC26673.1| unnamed protein product [Mus musculus]
HG1000175N0_160000_gene_prediction2	gi|26325838|dbj|BAC26673.1| unnamed protein product [Mus musculus]
HG1000176N0_1000_gene_prediction1	gi|26354216|dbj|BAC40736.1| unnamed protein product [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000176N0_160000_gene_prediction1	gi|26337635|dbj|BAC32503.1| unnamed protein product [Mus musculus]
HG1000177N0_160000_gene_prediction1	gi|26337635|dbj|BAC32503.1| unnamed protein product [Mus musculus]
HG1000178N0_160000_gene_prediction1	gi|20884040|ref|XP_134731.1| endothelial differentiation, sphingolipid G-protein-coupled receptor, 5 [Mus musculus]
HG1000178N0_160000_gene_prediction2	gi|13384830|ref|NP_079706.1| RIKEN cDNA 1110066C01 [Mus musculus]
HG1000180N0_1000_gene_prediction1	gi|13384830|ref|NP_079706.1| RIKEN cDNA 1110066C01 [Mus musculus]
HG1000181N0_10000_gene_prediction1	gi|13384730|ref|NP_079640.1| RIKEN cDNA 1110005A23 [Mus musculus]
HG1000181N0_160000_gene_prediction1	gi|25023031|ref|XP_205093.1| similar to hypothetical protein FLJ38281 [Homo sapiens] [Mus musculus]
HG1000183N0_160000_gene_prediction1	gi|26334755|dbj|BAC31078.1| unnamed protein product [Mus musculus]
HG1000186N0_20000_gene_prediction1	gi|27370150|ref|NP_766364.1| hypothetical protein D630002G06 [Mus musculus]
HG1000186N0_160000_gene_prediction2	
HG1000187N0_20000_gene_prediction1	gi|26342222|dbj|BAC34773.1| unnamed protein product [Mus musculus]
HG1000187N0_160000_gene_prediction3	
HG1000189N0_1000_gene_prediction1	gi|25024769|ref|XP_207136.1| similar to ORF2 [Mus musculus domesticus]
HG1000189N0_5000_gene_prediction1	gi|26325734|dbj|BAC26621.1| unnamed protein product [Mus musculus]
HG1000189N0_1000_gene_prediction2	gi|20879992|ref|XP_140210.1| similar to BG:DS01759.1 gene product [Drosophila melanogaster] [Mus musculus]
HG1000189N0_5000_gene_prediction2	gi|26325734|dbj|BAC26621.1| unnamed protein product [Mus musculus]
HG1000195N0_10000_gene_prediction1	gi|20879992|ref|XP_140210.1| similar to BG:DS01759.1 gene product [Drosophila melanogaster] [Mus musculus]
HG1000199N0_160000_gene_prediction1	gi|17390530|gb|AAH18231.1| Unknown (protein for MGC:19236) [Mus musculus]
HG1000201N0_10000_gene_prediction1	gi|20824845|ref|XP_131963.1| expressed sequence C77020 [Mus musculus]
HG1000203N0_5000_gene_prediction1	gi|27477269|ref|XP_209223.1| similar to Transforming protein RhoC (H9) [Homo

FP ID	Fantom Top Hit Annotation
	sapiens]
HG1000204N0_10000_gene_prediction 1	gi|26333233|dbj|BAC30334.1| unnamed protein product [Mus musculus]
HG1000209N0_160000_gene_predictio n2	gi|26326739|dbj|BAC27113.1| unnamed protein product [Mus musculus]
HG1000215N0_5000_gene_prediction1	gi|27369784|ref|NP_766142.1| hypothetical protein A230053P19 [Mus musculus]
HG1000215N0_1000_gene_prediction1	gi|6671756|ref|NP_031732.1| suppressor of cytokine signaling 2; cytokine inducible SH2-containing protein 2; high growth; STAT-induced STAT inhibitor 2; cytokine-inducible SH2 protein 2 [Mus musculus]
HG1000219N0_10000_gene_prediction 1	gi|26328915|dbj|BAC28196.1| unnamed protein product [Mus musculus]
HG1000221N0_160000_gene_predictio n1	gi|4504255|ref|NP_002097.1| H2A histone family, member Z; H2AZ histone [Homo sapiens]
HG1000221N0_20000_gene_prediction 1	gi|11360345|pir||T42725 actin binding protein ACF7, neural isoform 1 - mouse (fragment)
HG1000223N0_160000_gene_predictio n1	gi|11360345|pir||T42725 actin binding protein ACF7, neural isoform 1 - mouse (fragment)
HG1000225N0_160000_gene_predictio n1	gi|25019988|ref|XP_207469.1| similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1000235N0_160000_gene_predictio n1	gi|20137004|ref|NP_035320.1| proteasome (prosome, macropain) 28 subunit, beta; protease (prosome, macropain) 28 subunit, beta [Mus musculus]
HG1000236N0_160000_gene_predictio n1	gi|15617197|ref|NP_077135.1| ATPase, H+ transporting, lysosomal 13kD, V1 subunit G isoform 1; ATPase, H+ transporting, lysosomal (vacuolar proton pump) [Mus musculus]
HG1000238N0_160000_gene_predictio n1	gi|6671704|ref|NP_031664.1| chaperonin subunit 7 (eta) [Mus musculus]
HG1000238N0_5000_gene_prediction1	gi|6671549|ref|NP_031479.1| anti-oxidant protein 2; acidic calcium-independent phospholipase A2; peroxiredoxin 5; 1-Cys Prx [Mus musculus]
HG1000239N0_160000_gene_predictio n1	gi|6671549|ref|NP_031479.1| anti-oxidant protein 2; acidic calcium-independent phospholipase A2; peroxiredoxin 5; 1-Cys Prx [Mus musculus]
HG1000241N0_160000_gene_predictio n1	gi|7657357|ref|NP_056596.1| nucleosome assembly protein 1-like 1; nucleosome assembly protein-1 [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000243N0_160000_gene_prediction1	gi|4759158|ref|NP_004588.1| small nuclear ribonucleoprotein D2 polypeptide 16.5kDa; small nuclear ribonucleoprotein D2 polypeptide (16.5kD) [Homo sapiens]
HG1000243N0_160000_gene_prediction2	gi|8393534|ref|NP_058653.1| high mobility group protein 17 [Mus musculus]
HG1000245N0_1000_gene_prediction1	gi|8393534|ref|NP_058653.1| high mobility group protein 17 [Mus musculus]
HG1000250N0_160000_gene_prediction1	gi|12850132|dbj|BAB28604.1| unnamed protein product [Mus musculus]
HG1000252N0_160000_gene_prediction1	gi|20824845|ref|XP_131963.1| expressed sequence C77020 [Mus musculus]
HG1000255N0_10000_gene_prediction1	gi|17105394|ref|NP_000975.2| ribosomal protein L23a; 60S ribosomal protein L23a; melanoma differentiation-associated gene 20 [Homo sapiens]
HG1000262N0_160000_gene_prediction2	gi|13385532|ref|NP_080303.1| RIKEN cDNA 2700086I23 [Mus musculus]
HG1000263N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000264N0_5000_gene_prediction1	gi|26360198|dbj|BAB25612.2| unnamed protein product [Mus musculus]
HG1000264N0_5000_gene_prediction2	gi|21624617|ref|NP_081018.1| RIKEN cDNA 1110007M04 [Mus musculus]
HG1000265N0_160000_gene_prediction1	gi|21624617|ref|NP_081018.1| RIKEN cDNA 1110007M04 [Mus musculus]
HG1000266N0_0_gene_prediction1	gi|25070241|ref|XP_192786.1| proline rich protein expressed in brain [Mus musculus]
HG1000266N0_160000_gene_prediction1	gi|12584972|ref|NP_075021.1| lipin 3 [Mus musculus]
HG1000267N0_5000_gene_prediction1	gi|26340094|dbj|BAC33710.1| unnamed protein product [Mus musculus]
HG1000270N0_160000_gene_prediction1	gi|6679937|ref|NP_032110.1| glyceraldehyde-3-phosphate dehydrogenase [Mus musculus]
HG1000271N0_10000_gene_prediction1	gi|12844196|dbj|BAB26273.1| unnamed protein product [Mus musculus]
HG1000271N0_160000_gene_prediction1	gi|26345908|dbj|BAC36605.1| unnamed protein product [Mus musculus]
HG1000273N0_160000_gene_prediction1	gi|26345908|dbj|BAC36605.1| unnamed protein product [Mus musculus]
HG1000295N0_160000_gene_prediction1	gi|20888943|ref|XP_129258.1| cDNA sequence AF233884 [Mus musculus]
HG1000296N0_160000_gene_prediction1	gi|21313266|ref|NP_080089.1| RIKEN cDNA 1200003O06 [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000299N0_160000_gene_predictio n1	gi|25054735|ref|XP_192839.1| ATPas, class II, type 9B [Mus musculus]
HG1000300N0_10000_gene_prediction 1	gi|6753882|ref|NP_034349.1| FK506 binding protein 4 (59 kDa) [Mus musculus]
HG1000306N0_0_gene_prediction1	gi|25024769|ref|XP_207136.1| similar to ORF2 [Mus musculus domesticus]
HG1000306N0_0_gene_prediction2	
HG1000312N0_160000_gene_predictio n1	
HG1000314N0_1000_gene_prediction1	gi|4506283|ref|NP_003454.1| protein tyrosine phosphatase type IVA, member 1; Protein tyrosine phosphatase IVA1 [Homo sapiens]
HG1000315N0_160000_gene_predictio n1	gi|4506285|ref|NP_003470.1| protein tyrosine phosphatase type IVA, member 2, isoform 1; protein tyrosine phosphatase IVA; protein tyrosine phosphatase IVA2; phosphatase of regenerating liver 2 [Homo sapiens]
HG1000330N0_160000_gene_predictio n2	gi|6679553|ref|NP_033003.1| protein tyrosine phosphatase, non-receptor type 2 [Mus musculus]
HG1000330N0_160000_gene_predictio n4	gi|12860388|dbj|BAB31939.1| unnamed protein product [Mus musculus]
HG1000332N0_10000_gene_prediction 1	gi|26344091|dbj|BAC35702.1| unnamed protein product [Mus musculus]
HG1000337N0_1000_gene_prediction1	gi|20987322|gb|AAH30185.1| Unknown (protein for MGC:29401) [Mus musculus]
HG1000341N0_5000_gene_prediction1	gi|4506725|ref|NP_000998.1| ribosomal protein S4, X-linked X isoform; 40S ribosomal protein S4, X isoform; ribosomal protein S4X isoform; single-copy abundant mRNA; cell cycle gene 2 [Homo sapiens]
HG1000341N0_10000_gene_prediction 1	gi|26332837|dbj|BAC30136.1| unnamed protein product [Mus musculus]
HG1000353N0_160000_gene_predictio n1	gi|17157989|ref|NP_473384.1| Musashi homolog 2 (Drosophila) [Mus musculus]
HG1000357N0_20000_gene_prediction 1	gi|25021483|ref|XP_207941.1| similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1000358N0_5000_gene_prediction1	gi|27372319|dbj|BAC53724.1| Piccolo [Mus musculus]
HG1000359N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000363N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]

FP ID	Fantom Top Hit Annotation
HG1000364N0_160000_gene_predictio n1	gi|19484126|gb|AAH25846.1| Unknown (protein for MGC:32383) [Mus musculus]
HG1000367N0_160000_gene_predictio n1	gi|13928676|ref|NP_113687.1| proline rich protein 2 [Mus musculus]
HG1000379N0_160000_gene_predictio n1	gi|20863632|ref|XP_164160.1| hypothetical protein XP_164160 [Mus musculus]
HG1000390N0_10000_gene_prediction 1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000390N0_5000_gene_prediction1	gi|20892585|ref|XP_147977.1| RIKEN cDNA 2610001E17 [Mus musculus]
HG1000391N0_160000_gene_predictio n1	gi|20892585|ref|XP_147977.1| RIKEN cDNA 2610001E17 [Mus musculus]
HG1000396N0_160000_gene_predictio n2	gi|26330368|dbj|BAC28914.1| unnamed protein product [Mus musculus]
HG1000401N0_10000_gene_prediction 1	
HG1000407N0_160000_gene_predictio n1	gi|12853695|dbj|BAB29819.1| unnamed protein product [Mus musculus]
HG1000408N0_160000_gene_predictio n2	gi|25029560|ref|XP_203691.1| similar to PROBABLE POL POLYPROTEIN [Mus musculus]
HG1000414N0_160000_gene_predictio n2	gi|26326871|dbj|BAC27179.1| unnamed protein product [Mus musculus]
HG1000416N0_160000_gene_predictio n1	gi|20902061|ref|XP_147959.1| hypothetical protein XP_147959 [Mus musculus]
HG1000428N0_160000_gene_predictio n1	gi|25032567|ref|XP_207391.1| similar to ORF2 [Mus musculus domesticus]
HG1000429N0_160000_gene_predictio n1	gi|25022040|ref|XP_204233.1| similar to ORF2 [Mus musculus domesticus]
HG1000431N0_20000_gene_prediction 1	gi|26339864|dbj|BAC33595.1| unnamed protein product [Mus musculus]
HG1000435N0_160000_gene_predictio n1	gi|8394057|ref|NP_058565.1| low density lipoprotein receptor-related protein 4; low density lipoprotein-related protein 4; Low Density Lipoprotein Receptor Related Protein 4; corin [Mus musculus]
HG1000441N0_160000_gene_predictio n1	gi|26340972|dbj|BAC34148.1| unnamed protein product [Mus musculus]
HG1000441N0_160000_gene_predictio n2	gi|12836479|dbj|BAB23675.1| unnamed protein product [Mus musculus]
HG1000446N0_160000_gene_predictio n1	gi|25029827|ref|XP_207226.1| similar to ORF2 [Mus musculus domesticus]
HG1000446N0_160000_gene_predictio n2	gi|25031497|ref|XP_207552.1| similar to Retrovirus-related POL polyprotein [Mus

FP ID	Fantom Top Hit Annotation
	musculus]
HG1000449N0_160000_gene_predictio n2	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000451N0_160000_gene_predictio n1	gi|25054021|ref|XP_192811.1| similar to Transmembrane protease, serine 2 (Epitheliasin) (Plasmic transmembrane protein X) [Mus musculus]
HG1000455N0_10000_gene_prediction 1	gi|20846744|ref|XP_144090.1| similar to hypothetical protein FLJ12457 [Mus musculus]
HG1000461N0_10000_gene_prediction 1	gi|20824899|ref|XP_144255.1| hypothetical protein XP_144255 [Mus musculus]
HG1000474N0_5000_gene_prediction1	gi|12853695|dbj|BAB29819.1| unnamed protein product [Mus musculus]
HG1000476N0_1000_gene_prediction1	gi|12834707|dbj|BAB23011.1| unnamed protein product [Mus musculus]
HG1000489N0_160000_gene_predictio n1	no blast hit
HG1000499N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000500N0_160000_gene_predictio n1	gi|20912903|ref|XP_126663.1| RIKEN cDNA 2410154J16 [Mus musculus]
HG1000505N0_160000_gene_predictio n1	gi|25044951|ref|XP_195302.1| similar to olfactory receptor MOR256-23 [Mus musculus]
HG1000509N0_10000_gene_prediction 1	gi|26334721|dbj|BAC31061.1| unnamed protein product [Mus musculus]
HG1000510N0_160000_gene_predictio n1	gi|12834707|dbj|BAB23011.1| unnamed protein product [Mus musculus]
HG1000513N0_160000_gene_predictio n1	gi|12859663|dbj|BAB31727.1| unnamed protein product [Mus musculus]
HG1000519N0_160000_gene_predictio n1	gi|119146|sp|P20001|EF11_CRIGR Elongation factor 1-alpha 1 (EF-1-alpha-1) (Elongation factor 1 A-1) (eEF1A-1) (Elongation factor Tu) (EF-Tu)
HG1000521N0_160000_gene_predictio n1	gi|2495301|sp|Q63934|BR3B_MOUSE Brain-specific homeobox/POU domain protein 3B (BRN-3B) (BRN-3.2)
HG1000524N0_160000_gene_predictio n1	gi|21280325|dbj|BAB96760.1| type XXVI collagen [Mus musculus]
HG1000530N0_20000_gene_prediction 1	gi|6679921|ref|NP_032102.1| gamma-aminobutyric acid (GABA-A) receptor, subunit rho 2 [Mus musculus]
HG1000530N0_160000_gene_predictio n2	gi|23622684|ref|XP_156394.2| expressed sequence AL023001 [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000534N0_160000_gene_predictio n1	
HG1000545N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000549N0_160000_gene_predictio n1	gi|26341288|dbj|BAC34306.1| unnamed protein product [Mus musculus]
HG1000549N0_160000_gene_predictio n2	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000549N0_160000_gene_predictio n3	gi|21312126|ref|NP_081135.1| RIKEN cDNA 1110068E11 [Mus musculus]
HG1000553N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000560N0_160000_gene_predictio n2	gi|25032555|ref|XP_207412.1| similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1000562N0_160000_gene_predictio n1	
HG1000566N0_40000_gene_prediction 1	gi|20856064|ref|XP_151615.1| hypothetical protein XP_151615 [Mus musculus]
HG1000566N0_160000_gene_predictio n1	
HG1000582N0_160000_gene_predictio n1	gi|7656873|ref|NP_056579.1| RIKEN cDNA 5730583K22 gene [Mus musculus]
HG1000598N0_160000_gene_predictio n1	gi|4512261|dbj|BAA75227.1| neurochondrin-2 [Mus musculus]
HG1000606N0_20000_gene_prediction 1	gi|19527094|ref|NP_598640.1| expressed sequence AI327031 [Mus musculus]
HG1000607N0_160000_gene_predictio n1	gi|25058382|ref|XP_206318.1| hypothetical protein XP_206318 [Mus musculus]
HG1000608N0_20000_gene_prediction 1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000616N0_1000_gene_prediction1	gi|26387941|dbj|BAC25633.1| unnamed protein product [Mus musculus]
HG1000622N0_160000_gene_predictio n2	
HG1000623N0_160000_gene_predictio n1	gi|20904129|ref|XP_155605.1| hypothetical protein XP_155605 [Mus musculus]
HG1000624N0_160000_gene_predictio n1	gi|13542693|gb|AAH05553.1| putative chloride channel (similar to Mm Clcn4-2) [Mus musculus]
HG1000625N0_160000_gene_predictio n1	gi|20901495|ref|XP_140099.1| RIKEN cDNA 9130404H23 [Mus musculus]
HG1000628N0_40000_gene_prediction 1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]

FP ID	Fantom Top Hit Annotation
HG1000628N0_20000_gene_prediction 1	gi|26339720|dbj|BAC33523.1| unnamed protein product [Mus musculus]
HG1000638N0_5000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000642N0_160000_gene_predictio n1	
HG1000646N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000649N0_160000_gene_predictio n1	gi|25049717|ref|XP_149640.2| similar to gene Dbp73D protein - fruit fly (Drosophila melanogaster) [Mus musculus]
HG1000650N0_160000_gene_predictio n1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000652N0_160000_gene_predictio n2	gi|26377673|dbj|BAC25377.1| unnamed protein product [Mus musculus]
HG1000656N0_160000_gene_predictio n1	gi|13384666|ref|NP_079583.1| nuclear receptor binding factor 2 [Mus musculus]
HG1000656N0_160000_gene_predictio n2	gi|25050704|ref|XP_133465.2| RIKEN cDNA 2410004H02 [Mus musculus]
HG1000659N0_20000_gene_prediction 1	gi|25050704|ref|XP_133465.2| RIKEN cDNA 2410004H02 [Mus musculus]
HG1000661N0_20000_gene_prediction 1	gi|26333733|dbj|BAC30584.1| unnamed protein product [Mus musculus]
HG1000664N0_160000_gene_predictio n1	gi|27372319|dbj|BAC53724.1| Piccolo [Mus musculus]
HG1000670N0_160000_gene_predictio n1	gi|6680195|ref|NP_032255.1| histone deacetylase 2; DNA segment, Chr 10, Wayne State University 179, expressed [Mus musculus]
HG1000685N0_160000_gene_predictio n2	gi|17313266|ref|NP_478121.1| RecQ protein-like 4 [Mus musculus]
HG1000690N0_20000_gene_prediction 1	
HG1000690N0_20000_gene_prediction 2	gi|26340662|dbj|BAC33993.1| unnamed protein product [Mus musculus]
HG1000696N0_20000_gene_prediction 1	gi|26340662|dbj|BAC33993.1| unnamed protein product [Mus musculus]
HG1000696N0_40000_gene_prediction 1	gi|26326171|dbj|BAC26829.1| unnamed protein product [Mus musculus]
HG1000697N0_160000_gene_predictio n1	gi|25024387|ref|XP_207341.1| hypothetical protein XP_207341 [Mus musculus]
HG1000700N0_160000_gene_predictio n2	gi|26351279|dbj|BAC39276.1| unnamed protein product [Mus musculus]
HG1000704N0_160000_gene_predictio	gi|21644579|ref|NP_660253.1| Williams-

FP ID	Fantom Top Hit Annotation
n1	Beuren syndrome critical region gene 17 [Mus musculus]
HG1000711N0_20000_gene_prediction 1	gi|23273683|gb|AAH37239.1| Similar to BCL2-associated athanogene 4 [Mus musculus]
HG1000738N0_160000_gene_predictio n1	gi|12856848|dbj|BAB30802.1| unnamed protein product [Mus musculus]
HG1000739N0_160000_gene_predictio n1	gi|26339470|dbj|BAC33406.1| unnamed protein product [Mus musculus]
HG1000739N0_160000_gene_predictio n2	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000740N0_10000_gene_prediction 1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000743N0_160000_gene_predictio n1	gi|23601536|ref|XP_130965.2| Nice-4 protein homolog [Mus musculus]
HG1000779N0_160000_gene_predictio n1	gi|2627027|dbj|BAA23475.1| Ftp-1 [Mus musculus]
HG1000781N0_160000_gene_predictio n1	gi|25023334|ref|XP_204722.1| similar to formin [Mus musculus]
HG1000781N0_160000_gene_predictio n2	gi|26350877|dbj|BAC39075.1| unnamed protein product [Mus musculus]
HG1000786N0_160000_gene_predictio n1	gi|25023581|ref|XP_207103.1| similar to Retrovirus-related POL polyprotein [Mus musculus]
HG1000788N0_1000_gene_prediction1	gi|26340832|dbj|BAC34078.1| unnamed protein product [Mus musculus]
HG1000799N0_20000_gene_prediction 1	gi|20847912|ref|XP_144610.1| similar to KIAA1904 protein [Homo sapiens] [Mus musculus]
HG1000808N0_160000_gene_predictio n1	gi|26345960|dbj|BAC36631.1| unnamed protein product [Mus musculus]
HG1000817N0_160000_gene_predictio n1	gi|20882231|ref|XP_139203.1| similar to KIAA0858 protein [Homo sapiens] [Mus musculus]
HG1000822N0_20000_gene_prediction 1	gi|13242237|ref|NP_077327.1| Heat shock cognate protein 70; heat shock 70kD protein 8 [Rattus norvegicus]
HG1000824N0_160000_gene_predictio n1	gi|6680195|ref|NP_032255.1| histone deacetylase 2; DNA segment, Chr 10, Wayne State University 179, expressed [Mus musculus]
HG1000824N0_10000_gene_prediction 1	gi|20883564|ref|XP_152815.1| hypothetical protein XP_152815 [Mus musculus]
HG1000839N0_160000_gene_predictio n1	gi|20883564|ref|XP_152815.1| hypothetical protein XP_152815 [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1000842N0_160000_gene_prediction1	gi|26339496|dbj|BAC33419.1| unnamed protein product [Mus musculus]
HG1000842N0_160000_gene_prediction2	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000869N0_160000_gene_prediction1	gi|6715564|ref|NP_032607.1| melanoma antigen, 80 kDa [Mus musculus]
HG1000870N0_160000_gene_prediction1	gi|20881174|ref|XP_147875.1| hypothetical protein XP_147875 [Mus musculus]
HG1000870N0_160000_gene_prediction2	gi|27369942|ref|NP_766246.1| hypothetical protein 9530051F04 [Mus musculus]
HG1000878N0_20000_gene_prediction1	gi|27369942|ref|NP_766246.1| hypothetical protein 9530051F04 [Mus musculus]
HG1000878N0_20000_gene_prediction2	gi|27369942|ref|NP_766246.1| hypothetical protein 9530051F04 [Mus musculus]
HG1000904N0_160000_gene_prediction2	gi|27369942|ref|NP_766246.1| hypothetical protein 9530051F04 [Mus musculus]
HG1000904N0_40000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000906N0_5000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000906N0_160000_gene_prediction2	gi|20836822|ref|XP_130277.1| similar to Plakophilin 4 (p0071) [Mus musculus]
HG1000910N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000948N0_160000_gene_prediction1	gi|26325846|dbj|BAC26677.1| unnamed protein product [Mus musculus]
HG1000955N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1000959N0_160000_gene_prediction1	gi|7670427|dbj|BAA95065.1| unnamed protein product [Mus musculus]
HG1000959N0_5000_gene_prediction1	gi|22507385|ref|NP_081019.1| RIKEN cDNA 1110014F12 [Mus musculus]
HG1000990N0_5000_gene_prediction1	gi|22507385|ref|NP_081019.1| RIKEN cDNA 1110014F12 [Mus musculus]
HG1000994N0_10000_gene_prediction1	gi|10946762|ref|NP_067382.1| triggering receptor expressed on myeloid cells 3; triggering receptor expressed on monocytes 3 [Mus musculus]
HG1000994N0_160000_gene_prediction2	gi|12855175|dbj|BAB30238.1| unnamed protein product [Mus musculus]
HG1000994N0_10000_gene_prediction2	gi|12855175|dbj|BAB30238.1| unnamed protein product [Mus musculus]
HG1001001N0_160000_gene_prediction1	gi|12855175|dbj|BAB30238.1| unnamed protein product [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1001001N0_0_gene_prediction1	gi|26337385|dbj|BAC32378.1| unnamed protein product [Mus musculus]
HG1001002N0_160000_gene_predictio n1	gi|27370034|ref|NP_766297.1| hypothetical protein A530025J20 [Mus musculus]
HG1001003N0_0_gene_prediction1	gi|20348159|ref|XP_111588.1| similar to TRAV9D-3 [Mus musculus]
HG1001007N0_160000_gene_predictio n2	gi|27370034|ref|NP_766297.1| hypothetical protein A530025J20 [Mus musculus]
HG1001011N0_160000_gene_predictio n1	gi|13097000|gb|AAH03291.1| Similar to hypothetical protein FLJ10342 [Mus musculus]
HG1001011N0_160000_gene_predictio n2	gi|26336525|dbj|BAC31945.1| unnamed protein product [Mus musculus]
HG1001014N0_160000_gene_predictio n1	gi|25047957|ref|XP_130582.2| similar to hypothetical protein MGC14161 [Homo sapiens] [Mus musculus]
HG1001014N0_5000_gene_prediction1	gi|26337385|dbj|BAC32378.1| unnamed protein product [Mus musculus]
HG1001017N0_160000_gene_predictio n1	gi|26337385|dbj|BAC32378.1| unnamed protein product [Mus musculus]
HG1001020N0_160000_gene_predictio n1	gi|25019831|ref|XP_207463.1| similar to CD59B [Mus musculus]
HG1001024N0_160000_gene_predictio n1	gi|26338976|dbj|BAC33159.1| unnamed protein product [Mus musculus]
HG1001024N0_160000_gene_predictio n2	gi|20915148|ref|XP_149841.1| hypothetical protein XP_149841 [Mus musculus]
HG1001031N0_160000_gene_predictio n1	gi|20915148|ref|XP_149841.1| hypothetical protein XP_149841 [Mus musculus]
HG1001035N0_5000_gene_prediction1	gi|25071690|ref|XP_193591.1| hypothetical protein XP_193591 [Mus musculus]
HG1001043N0_160000_gene_predictio n1	gi|26347249|dbj|BAC37273.1| unnamed protein product [Mus musculus]
HG1001046N0_5000_gene_prediction1	gi|6678714|ref|NP_032537.1| lymphoid-restricted membrane protein [Mus musculus]
HG1001046N0_160000_gene_predictio n1	gi|25048969|ref|XP_143803.3| similar to bA4O1.1 (novel protein) [Homo sapiens] [Mus musculus]
HG1001047N0_1000_gene_prediction1	gi|25021180|ref|XP_207917.1| similar to RNP particle component [Mus musculus]
HG1001048N0_160000_gene_predictio n1	gi|26353724|dbj|BAC40492.1| unnamed protein product [Mus musculus]
HG1001048N0_160000_gene_predictio n2	gi|20343845|ref|XP_109652.1| similar to hypothetical protein FLJ25217 [Homo sapiens] [Mus musculus]
HG1001144N0_20000_gene_prediction	gi|20346197|ref|XP_110161.1| RAN binding

FP ID	Fantom Top Hit Annotation
1	protein 1 [Mus musculus]
HG1001148N0_160000_gene_prediction2	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1001172N0_160000_gene_prediction1	gi|26339628|dbj|BAC33485.1| unnamed protein product [Mus musculus]
HG1001172N0_20000_gene_prediction1	gi|22122489|ref|NP_666128.1| hypothetical protein MGC38936 [Mus musculus]
HG1001187N0_160000_gene_prediction1	gi|26340706|dbj|BAC34015.1| unnamed protein product [Mus musculus]
HG1001192N0_160000_gene_prediction1	gi|18497290|ref|NP_084056.1| protein kinase raf 1; murine sarcoma 3611 oncogene 1; sarcoma 3611 oncogene [Mus musculus]
HG1001194N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1001199N0_160000_gene_prediction1	gi|20837732|ref|XP_132241.1| hypothetical protein XP_132241 [Mus musculus]
HG1001199N0_160000_gene_prediction2	gi|20071068|gb|AAH27341.1| Similar to elongation factor G2 [Mus musculus]
HG1001220N0_160000_gene_prediction1	gi|20071068|gb|AAH27341.1| Similar to elongation factor G2 [Mus musculus]
HG1001223N0_160000_gene_prediction1	gi|20908735|ref|XP_122598.1| similar to helix-destabilizing protein - rat [Mus musculus]
HG1001229N0_160000_gene_prediction2	gi|25024769|ref|XP_207136.1| similar to ORF2 [Mus musculus domesticus]
HG1001230N0_5000_gene_prediction1	gi|6754206|ref|NP_034568.1| hexokinase 1; downeast anemia [Mus musculus]
HG1001235N0_160000_gene_prediction1	gi|12857205|dbj|BAB30930.1| unnamed protein product [Mus musculus]
HG1001235N0_10000_gene_prediction1	gi|21703918|ref|NP_663438.1| hypothetical protein BC024118 [Mus musculus]
HG1001235N0_20000_gene_prediction1	gi|26339338|dbj|BAC33340.1| unnamed protein product [Mus musculus]
HG1001235N0_160000_gene_prediction2	gi|26339338|dbj|BAC33340.1| unnamed protein product [Mus musculus]
HG1001235N0_160000_gene_prediction3	gi|26340904|dbj|BAC34114.1| unnamed protein product [Mus musculus]
HG1001260N0_160000_gene_prediction1	gi|26327795|dbj|BAC27638.1| unnamed protein product [Mus musculus]
HG1001260N0_40000_gene_prediction1	gi|8922328|ref|NP_060517.1| hypothetical protein FLJ10290 [Homo sapiens]
HG1001264N0_160000_gene_prediction1	gi|8922328|ref|NP_060517.1| hypothetical protein FLJ10290 [Homo sapiens]
HG1001274N0_160000_gene_predictio	gi|26383198|dbj|BAC25520.1| unnamed protein

FP ID	Fantom Top Hit Annotation
n1	product [Mus musculus]
HG1001284N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1001284N0_160000_gene_prediction2	gi|26326843|dbj|BAC27165.1| unnamed protein product [Mus musculus]
HG1001292N0_160000_gene_prediction1	gi|26326843|dbj|BAC27165.1| unnamed protein product [Mus musculus]
HG1001302N0_160000_gene_prediction1	gi|13097342|gb|AAH03421.1| Similar to ATPase, H+ transporting, lysosomal (vacuolar proton pump) 31kD [Mus musculus]
HG1001313N0_160000_gene_prediction1	gi|12852631|dbj|BAB29486.1| unnamed protein product [Mus musculus]
HG1001323N0_160000_gene_prediction1	gi|25053141|ref|XP_193739.1| similar to betaine-homocysteine methyltransferase [Rattus norvegicus] [Mus musculus]
HG1001328N0_5000_gene_prediction1	gi|26347687|dbj|BAC37492.1| unnamed protein product [Mus musculus]
HG1001328N0_40000_gene_prediction1	gi|26352918|dbj|BAC40089.1| unnamed protein product [Mus musculus]
HG1001331N0_0_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1001335N0_160000_gene_prediction1	gi|20381292|gb|AAH27770.1| stromal cell derived factor receptor 2 [Mus musculus]
HG1001335N0_160000_gene_prediction2	gi|2193870|dbj|BAA20419.1| reverse transcriptase [Mus musculus]
HG1001348N0_160000_gene_prediction1	gi|2193870|dbj|BAA20419.1| reverse transcriptase [Mus musculus]
HG1001349N0_160000_gene_prediction1	gi|20846538|ref|XP_150033.1| hypothetical protein XP_150033 [Mus musculus]
HG1001354N0_160000_gene_prediction1	gi|7305215|ref|NP_038599.1| kinase suppressor of ras [Mus musculus]
HG1001361N0_160000_gene_prediction1	gi|6678690|ref|NP_032525.1| LIM homeobox protein 5; LIM homeo box protein 5 [Mus musculus]
HG1001376N0_160000_gene_prediction1	gi|20345901|ref|XP_109824.1| hypothetical protein XP_109824 [Mus musculus]
HG1001376N0_5000_gene_prediction1	gi|27261816|ref|NP_080861.1| RIKEN cDNA C530005J20 [Mus musculus]
HG1001376N0_20000_gene_prediction1	gi|27261816|ref|NP_080861.1| RIKEN cDNA C530005J20 [Mus musculus]
HG1001376N0_5000_gene_prediction2	gi|27261816|ref|NP_080861.1| RIKEN cDNA C530005J20 [Mus musculus]
HG1001376N0_5000_gene_prediction3	gi|27261816|ref|NP_080861.1| RIKEN cDNA C530005J20 [Mus musculus]

FP ID	Fantom Top Hit Annotation
HG1001417N0_160000_gene_prediction1	gi|27261816|ref|NP_080861.1| RIKEN cDNA C530005J20 [Mus musculus]
HG1001417N0_1000_gene_prediction1	gi|26349767|dbj|BAC38523.1| unnamed protein product [Mus musculus]
HG1001417N0_160000_gene_prediction2	gi|26349767|dbj|BAC38523.1| unnamed protein product [Mus musculus]
HG1001417N0_160000_gene_prediction3	gi|26349767|dbj|BAC38523.1| unnamed protein product [Mus musculus]
HG1001436N0_5000_gene_prediction1	gi|26349767|dbj|BAC38523.1| unnamed protein product [Mus musculus]
HG1001436N0_20000_gene_prediction1	gi|20987280|gb|AAH29643.1| Unknown (protein for MGC:25768) [Mus musculus]
HG1001436N0_160000_gene_prediction1	gi|25051637|ref|XP_194491.1| RIKEN cDNA 1110053F02 [Mus musculus]
HG1001439N0_160000_gene_prediction1	gi|25051637|ref|XP_194491.1| RIKEN cDNA 1110053F02 [Mus musculus]
HG1001484N0_160000_gene_prediction1	gi|6753290|ref|NP_033943.1| calsequestrin 1 [Mus musculus]
HG1001485N0_10000_gene_prediction1	gi|25029827|ref|XP_207226.1| similar to ORF2 [Mus musculus domesticus]
HG1001500N0_160000_gene_prediction1	gi|3599320|gb|AAC72793.1| ORF2 [Mus musculus domesticus]
HG1001500N0_160000_gene_prediction2	gi|6679108|ref|NP_032748.1| nucleophosmin 1; nucleolar protein NO38 [Mus musculus]
HG1001508N0_160000_gene_prediction1	gi|25029928|ref|XP_207257.1| similar to Retrovirus-related POL polyprotein [Mus musculus]
	gi|20340683|ref|XP_110361.1| similar to phospholipase C beta 2 [Rattus norvegicus] [Mus musculus]

Examples

[0423] The examples, which are intended to be purely exemplary of the invention and should therefore not be considered to limit the invention in any way, also describe and detail aspects and embodiments of the invention discussed above. The examples are not intended to represent that the experiments below are all or the only experiments performed. Efforts have been made to ensure accuracy with respect to numbers used (e.g., amounts, temperature, etc.) but some experimental errors and deviations should be accounted for. Unless indicated otherwise, parts are parts by weight, molecular weight is weight average molecular weight, temperature is in degrees Centigrade, and pressure is at or near atmospheric.

[0424] While the present invention has been described with reference to the specific embodiments thereof, it should be understood by those skilled in the art that various changes may be made and equivalents may be substituted without departing from the true spirit and scope of the invention. In addition, many modifications can be made to adapt a particular situation, material, composition of matter, process, process step or steps, to the objective, spirit and scope of the present invention. All such modifications are intended to be within the scope of the claims appended hereto.

[0425] Additional objects and advantages of the invention will be set forth in part in the description which follows, and in part will be obvious from the description, or may be learned by practice of the invention. The objects and advantages of the invention will be realized and attained by means of the elements and combinations particularly pointed out in the appended claims. Moreover, advantages described in the body of the specification, if not included in the claims, are not per se limitations to the claimed invention.

[0426] It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only and are not restrictive of the invention, as claimed. Moreover, it must be understood that the invention is not limited to the particular embodiments described, as such may, of course, vary. Further, the terminology used to describe particular embodiments is not intended to be limiting, since the scope of the present invention will be limited only by its claims.

[0427] With respect to ranges of values, the invention encompasses each intervening value between the upper and lower limits of the range to at least a tenth of the lower limit's unit, unless the context clearly indicates otherwise. Further, the

invention encompasses any other stated intervening values. Moreover, the invention also encompasses ranges excluding either or both of the upper and lower limits of the range, unless specifically excluded from the stated range.

[0428] Unless defined otherwise, the meanings of all technical and scientific terms used herein are those commonly understood by one of ordinary skill in the art to which this invention belongs. One of ordinary skill in the art will also appreciate that any methods and materials similar or equivalent to those described herein can also be used to practice or test the invention. Further, all publications mentioned herein are incorporated by reference.

[0429] It must be noted that, as used herein and in the appended claims, the singular forms "a," "or," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a subject polypeptide" includes a plurality of such polypeptides and reference to "the agent" includes reference to one or more agents and equivalents thereof known to those skilled in the art, and so forth.

[0430] Further, all numbers expressing quantities of ingredients, reaction conditions, % purity, polypeptide and polynucleotide lengths, and so forth, used in the specification and claims, are modified by the term "about," unless otherwise indicated. Accordingly, the numerical parameters set forth in the specification and claims are approximations that may vary depending upon the desired properties of the present invention. At the very least, and not as an attempt to limit the application of the doctrine of equivalents to the scope of the claims, each numerical parameter should at least be construed in light of the number of reported significant digits, applying ordinary rounding techniques. Nonetheless, the numerical values set forth in the specific examples are reported as precisely as possible. Any numerical value, however, inherently contains certain errors from the standard deviation of its experimental measurement.

[0431] The publications discussed herein are provided solely for their disclosure prior to the filing date of the present application. Nothing herein is to be construed as an admission that the present invention is not entitled to antedate such publication by virtue of prior invention. Further, the dates of publication provided may be different from the actual publication dates which may need to be independently confirmed.

Example 1 Expression in *E. coli*

[0432] Sequences can be expressed in *E. coli*. Any one or more of the sequences according to SEQ ID NOS.: 1 -209 and 419 - 627 can be expressed in *E. coli* by subcloning the entire coding region, or a selected portion thereof, into a prokaryotic expression vector. For example, the expression vector pQE16 from the QIA expression prokaryotic protein expression system (Qiagen, Valencia, CA) can be used. The features of this vector that make it useful for protein expression include an efficient promoter (phage T5) to drive transcription, expression control provided by the lac operator system, which can be induced by addition of IPTG (isopropyl-beta-D-thiogalactopyranoside), and an encoded 6XHis tag coding sequence. The latter is a stretch of six histidine amino acid residues which can bind very tightly to a nickel atom. This vector can be used to express a recombinant protein with a 6XHis. tag fused to its carboxyl terminus, allowing rapid and efficient purification using Ni-coupled affinity columns.

[0433] The entire or the selected partial coding region can be amplified by PCR, then ligated into digested pQE16 vector. The ligation product can be transformed by electroporation into electrocompetent *E. coli* cells (for example, strain M15[pREP4] from Qiagen), and the transformed cells may be plated on ampicillin-containing plates. Colonies may then be screened for the correct insert in the proper orientation using a PCR reaction employing a gene-specific primer and a vector-specific primer. Also, positive clones can be sequenced to ensure correct orientation and sequence. To express the proteins, a colony containing a correct recombinant clone can be inoculated into L-Broth containing 100 μg/ml of ampicillin, and 25 μg/ml of kanamycin, and the culture allowed to grow overnight at 37 degrees C. The saturated culture may then be diluted 20-fold in the same medium and allowed to grow to an optical density of 0.5 at 600 nm. At this point, IPTG can be added to a final concentration of 1 mM to induce protein expression. After growing the culture for an additional 5 hours, the cells may be harvested by centrifugation at 3000 times *g* for 15 minutes.

[0434] The resultant pellet can be lysed with a mild, nonionic detergent in 20 mM Tris HCl (pH 7.5) (B PER.TM. Reagent from Pierce, Rockford, IL), or by sonication until the turbid cell suspension turns translucent. The resulting lysate can be further purified using a nickel-containing column (Ni-NTA spin column from

Qiagen) under non-denaturing conditions. Briefly, the lysate will be adjusted to 300 mM NaCl and 10 mM imidazole, then centrifuged at 700 times *g* through the nickel spin column to allow the His-tagged recombinant protein to bind to the column. The column will be washed twice with wash buffer (for example, 50 mM NaH_2PO_4, pH 8.0; 300 mM NaCl; 20 mM imidazole) and eluted with elution buffer (for example, 50 mM NaH2 PO4, pH 8.0; 300 mM NaCl; 250 mM imidazole). All the above procedures will be performed at 4 degrees C. The presence of a purified protein of the predicted size can be confirmed with SDS-PAGE.

Example 2: Expression in Mammalian Cells

[0435] The sequences encoding the proteins of Example 1 can be cloned into the pENTR vector (Invitrogen) by PCR and transferred to the mammalian expression vector pDEST12.2 per manufacturer's instructions (Invitrogen). Introduction of the recombinant construct into the host cell can be effected by transfection with Fugene 6 (Roche) per manufacturer's instructions. The host cells containing one of polynucleotides of the invention can be used in conventional manners to produce the gene product encoded by the isolated fragment (in the case of an ORF). A number of types of cells can act as suitable host cells for expression of the proteins. Mammalian host cells include, for example, monkey COS cells, Chinese Hamster Ovary (CHO) cells, human kidney 293 cells, human epidermal A431 cells, human Colo205 cells, 3T3 cells, CV-1 cells, other transformed primate cell lines, normal diploid cells, cell strains derived from *in vitro* culture of primary tissue, primary explants, HeLa cells, mouse L cells, BHK, HL-60, U937, HaK or Jurkat cells.

Example 3: Expression in Cell-Free Translation Systems

[0436] Cell-free translation systems can also be employed to produce proteins using RNAs derived from the DNA constructs of the present invention. Appropriate cloning and expression vectors containing SP6 or T7 promoters for use with prokaryotic and eukaryotic hosts have been described (Sambrook et al., 1989). These DNA constructs can be used to produce proteins in a rabbit reticulocyte lysate system or in a wheat germ extract system.

[0437] Specific expression systems of interest include plant, bacterial, yeast, insect cell and mammalian cell derived expression systems. Expression systems in plants include those described in U.S. Patent No. 6,096,546 and U.S. Patent No. 6,127,145. Expression systems in bacteria include those described by Chang et al.,

1978, Goeddel et al., 1979, Goeddel et al., 1980, EP 0 036,776, U.S. Patent No. 4,551,433; DeBoer et al., 1983, and Siebenlist et al., 1980.

[0438] Mammalian expression is further accomplished as described in Dijkema et al. 1985, Gorman et al., 1982, Boshart et al., 1985, and U.S. Patent No. 4,399,216. Other features of mammalian expression are facilitated as described in Ham and Wallace, Meth. Enz., 1979, Barnes and Sato, 1980, U.S. Patent Nos. 4,767,704, 4,657,866, 4,927,762, 4,560,655, WO 90/103430, WO 87/00195, and U.S. RE 30,985.

Example 4: Expression of the Secreted Factors in Yeast

[0439] Primers can be designed to amplify the secreted factors using PCR and cloned into pENTR/D-TOPO vectors (Invitrogen, Carlsbad, CA). The secreted factors in pENTR/D-TOPO can be cloned into the yeast expression vector pYES-DEST52 by Gateway LR reaction (Invitrogen, Carlsbad, CA). The resulting yeast expression vectors can be transformed into INVSc1 strain from Invitrogen to express the secreted factors according to the manufacturer's protocol (Invitrogen, Carlsbad CA). The expressed secreted factors will have a 6XHis tag at the C-terminal. Expressed protein can be purified with ProBond™ resin (Invitrogen, Carlsbad, CA).

[0440] Expression systems in yeast include those described in Hinnen et al., 1978, Ito et al., 1983, Kurtz et al., 1986, Kunze et al., 1985, Gleeson et al., 1986, Roggenkamp et al., 1986, Das et al., 1984, De Louvencourt et al., 1983, Van den Berg et al., 1990, Kunze et al., 1985, Cregg et al. 1985, U.S. Patent No. 4,837,148, U.S. Patent No. 4,929,555, Beach and Nurse, 1981, Davidow et al., 1985, Gaillardin et al., 1985, Ballance et al., 1983, Tilburn et al., 1983, Yelton et al., 1984, Kelly and Hynes, 1985, EP 0 244,234, and WO 91/00357.

Example 5: Expression of Secreted Factors in Baculovirus Expression System.

[0441] The secreted factors in pENTR/D-TOPO can be cloned into Baculovirus expression vector pDEST10 by Gateway LR reaction (Invitrogen, Carlsbad, CA). The secreted factors can be expressed by the Bac-to-Bac expression system from Invitrogen (Carlsbad CA), briefly described as follows. The expression vectors containing the secreted factors are transformed into competent DH10Bac™ *E. coli* strain and selected for transposition. The resulting *E coli* contain recombinant bacmid that contains the secreted factor. High molecular weight DNA can be isolated from the *E. coli* containing the recombinant bacmid and then transfected into insect

cells with Cellfectin reagent. The expressed secreted factors will have a 6XHis tag at N-terminal. Expressed protein will be purified by ProBond™ resin (Invitrogen, Carlsbad, CA).

[0442] Expression of heterologous genes in insects can be accomplished as described in U.S. Patent No. 4,745,051; Doerfler et al., 1087; Friesen *et al.*, 1986; EP 0 127,839, EP 0 155,476, Vlak *et al.*, 1988, Miller *et al.*, 1988, Carbonell *et al.*, 1988, Maeda *et al.*, 1985, Lebacq-Verheyden *et al.*, 1988, Smith *et al.*, 1985, Miyajima *et al.*; and Martin *et al.*, 1988. Numerous baculoviral strains and variants and corresponding permissive insect host cells from hosts have been previously described (Setlow et al., 1986, Luckow *et al.*, 1988; Miller *et al.*, 1986; Maeda *et al.*, 1985).

Example 6: Primer Design

[0443] To design the forward primer for PCR amplification, the melting point of the first 20 to 24 bases of the primer can be calculated by counting total A and T residues, then multiplying by 2. To design the reverse primer for PCR amplification, the melting point of the first 20 to 24 bases of the reverse complement, with the sequences written from 5-prime to 3-prime can be calculated by counting the total G and C residues, then multiplying by 4. Both start and stop codons can be present in the final amplified clone. The length of the primers is such to obtain melting temperatures within 63 degrees C to 68 degrees C. Adding the bases "CACC" to the forward primer renders it compatible for cloning the PCR product with the TOPO pENTR/D (Invitrogen, CA).

Example 7: Reverse Transcriptase Reaction

[0444] cDNA can be prepared by the following method. Between 200 ng and 1.0 µg mRNA is added to 2 µl DMSO and the volume adjusted to 11 µl with DEPC-treated water. One µl Oligo dT is added to the tube, and the mixture is heated at 70° C for 5 min., quickly chilled on ice for 2 min., and the mixture is collected at the bottom of the tube by brief centrifugation. The following 1st strand components are then added to the mRNA mixture: 2 µl 10X Stratascript (Stratagene, CA) 1st strand buffer, 1 µl 0.1 M DTT, 1 µl 10 mM dNTP mix (10 mM each of dG, dA, dT and dCTP), 1 µl RNAse inhibitor, 3 µl Stratascript RT (50 U/ µl). The contents are gently mixed and the mixture collected by brief centrifugation. The mixture is incubated in a 42° C water bath for 1 hour, placed in a 70° C water bath for 15 min. to stop the reaction, transferred to ice for 2 min., and centrifuged briefly in a microfuge to collect the reaction product at the bottom of the reaction vessel. Two µl RNAse H is then

added to the tube, the contents are mixed well, incubated at 37° C in a water bath for 20 min., and centrifuged briefly in a microfuge to collect the reaction product at the bottom of the reaction vessel. The reaction mixture can proceed directly to PCR or be stored at – 20° C.

Example 8: Full Length PCR

[0445] Full length PCR can be achieved by placing the products of the reaction described in Example 7, with primers diluted to 5μM in water, into a reaction vessel and adding a reaction mixture composed of 1x Taq buffer, 25 mM dNTP, 10 ng cDNA pool, TaqPlus (Stratagene, CA) (5u/ul), PfuTurbo (Stratagene, CA) (2.5u/ul), water. The contents of the reaction vessel are then mixed gently by inversion 5-6 times, placed into a reservoir where 2μl F_1/R_1 primers are added, the plate sealed and placed in the thermocycler. The PCR reaction is comprised of the following eight steps. Step 1: 95° C for 3 min. Step 2: 94° C for 45 sec. Step 3: 0.5° C/sec to 56-60° C. Step 4: 56-60° C for 50 sec. Step 5: 72° C for 5 min. Step 6: Go to step 2, perform 35-40 cycles. Step 7: 72° C for 20 min. Step 8: 4° C.

[0446] The products can then be separated on a standard 0.8 to 1.0% agarose gel at 40 to 80 V, the bands of interest excised by cutting from the gel, and stored at – 20° C until extraction. The material in the bands of interest can be purified with QIAquick 96 PCR Purification Kit (Qiagen, CA) according to the manufacturer instructions. Cloning can be performed with the Topo Vector pENTR/D-TOPO vector (Invitrogen, CA) according to the manufacturer's instructions.

References

[0447] The specification is most thoroughly understood in light of the following references, all of which are hereby incorporated by reference in their entireties. The disclosures of the patents and other references cited above are also hereby incorporated by reference.

1. Agou, F., Quevillon, S., Kerjan, P., Latreille, M.T., Mirande, M. (1996) Functional replacement of hamster lysyl-tRNA synthetase by the yeast enzyme requires cognate amino acid sequences for proper tRNA recognition. *Biochemistry* 35:15322-15331.
2. Agrawal, S., Crooke, S.T. eds. (1998) Antisense Research and Application (Handbook of Experimental Pharmacology, Vol 131). Springer-Verlag New York, Inc.
3. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.D. (1994) Molecular Biology of the Cell. 3rd ed. Garland Publishing, Inc.
4. Alexander, D.R. (2000) The CD45 tyrosine phosphatase: a positive and negative regulator of immune cell function. *Semin. Immunol* 12:349-359.
5. Allison, A.C. (2000) Immunosuppressive drugs: the first 50 years and a glance forward. *Immunopharmacology* 47:63-83.
6. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. (1990) Basic alignment search tool. *J. Mol. Biol.* 215:403-410.
7. Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zheng, Z., Miller, W., Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* 25:3389-3402.
8. Amor, J.C., Harrison, D.H., Kahn, R.A., Ringe, D. (1994) Structure of the human ADP-ribosylation factor 1 complexed with GDP. *Nature* 372:704-708.
9. Andreeff, M., Pinkel, D. eds. (1999) Introduction to Fluorescence In Situ Hybridization: Principles and Clinical Applications. John Wiley & Sons.
10. Andres, D.A., Shao, H., Crick, D.C., Finlin, B.S. (1997) Expression cloning of a novel farnesylated protein, RDJ2, encoding a DnaJ protein homologue. *Arch. Biochem. Biophys.* 346:113-124.
11. Aubry, M., Marineau, C., Zhang, F.R., Zahed, L., Figlewicz, D., Delattre, O., Thomas, G., de Jong, P.J., Julien, J.P., Rouleau, G.A. (1992) Cloning of six new genes with zinc finger motifs mapping to short and long arms of human acrocentric chromosome 22 (p and q11.2). *Genomics* 13:641-648.

12. Ausubel, F., Brent. R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., eds. (1999) Short Protocols in Molecular Biology. 4th ed. Wiley & Sons.

13. Baksh, S., Burakoff, S.J. (2000) The role of calcineurin in lymphocyte activation. *Semin. Immunol.* 12:405-415.

14. Ballance, D.J., Buxton, F.P., Turner, G. (1983) Transformation of *Aspergillus nidulans* by the orotidine-5'-phosphate decarboxylase gene of *Neurospora crassa*. *Biochem. Biophys. Res. Commun.* 112:284-289.

15. Barnes, D., Sato, G. (1980) Methods for growth of cultured cells in serum-free medium. *Anal. Biochem.* 102:255-270.

16. Bashkin, J.K., Sampath, U., Frolova, E. (1995) Ribozyme mimics as catalytic antisense reagents. *Appl. Biochem. Biotechnol.* 54:43-56.

17. Bassett, D.E., Eisen, M.B., Boguski, M.S. (1999) Gene expression informatics - it's all in your mine. *Nature Genetics* 21:51-55.

18. Bast, R.C., Kufe, D.W., Pollock, R.E., Weichselbaum, R.R., Holland, J.F., Frei, E., eds. (2000) Cancer Medicine. 5th ed. B.C. Decker, Inc.

19. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M., Sonnhammer, E.L.L. (2000) *Nucleic Acids Research* 30:276-280.

20. Battini, R., Ferrari, S., Kaczmarek, L., Calabretta, B., Chen, S.T., Baserga, R. (1987) Molecular cloning of a cDNA for a human ADP/ATP carrier which is growth-regulated. *J. Biol. Chem.* 262:4355-4359.

21. Beach, D., Durkacz, B., Nurse, P. (1982) Functionally homologous cell cycle control genes in budding and fission yeast. *Nature* 300:706-709.

22. Bennett, J. (2000) Gene therapy for retinitis pigmentosa. *Curr. Opin. Mol. Ther.* 2:420-425.

23. Berinstein, N.L. (2002) Carcinoembryonic antigen as a target for therapeutic anticancer vaccines: a review. *J. Clin. Oncol.* 20:2197-2207.

24. Bibikova, M., Beumer, K., Trautman, J.K., Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* 300:764.

25. Birney, E., Durbin, R. (2000) Using GeneWise in the *Drosophila* annotation experiment. *Genome Res.* 10:547-548.

26. Bodzioch, M., Orso, E., Klucken, J., Langmann, T., Bottcher, A., Diederich, W., Drobnik, W., Barlage, S., Buchler, C., Porsch-Ozcurumez, M., Kaminski,

W.E., Hahmann, H.W., Oette, K., Rothe, G., Aslanidis, C., Lackner, K.J., Schmitz, G. (1999) The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. *Nat. Genet.* 1999 22:347-351.

27. Bonifaci, N., Moroianu, J., Radu, A., Blobel, G. (1997) Karyopherin beta2 mediates nuclear import of a mRNA binding protein. *Proc. Natl. Acad. Sci.* 94:5055-5060.
28. Bono, H., Kasukawa, T., Furuno, M., Hayashizaki, Y., Okazaki, Y. (2002) FANTOM DB: database of Functional Annotation of RIKEN Mouse cDNA Clones. *Nucleic Acids Res.* 30:116-118.
29. Boshart, M., Weber, F., Jahn, G., Dorsch-Hasler, K., Fleckenstein, B., Schaffner, W. (1985) A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* 41:521-530.
30. Bowtell, D.D.L. (1999) Options available - from start to finish - for obtaining expression data by microarray. *Nature Genetics* 21:25-32.
31. Brenner, S., Williams, S.R., Vermass, E.H., Storck, T., Moon, K., McCollum, C., Mao, J.I., Luo, S., Kirchner, J.J., Eletr, S., DuBridge, R.B., Burcham, T., Albrecht, G. (2000) *In vitro* cloning of complex mixtures of DNA on microbeads: physical separation of differentially expressed cDNAs. *Proc. Natl. Acad. Sci. USA* 97:1665-1670.
32. Brock, G. (2000) Sildenafil citrate (Viagra®). *Drugs Today* 36:125-134.
33. Brown, J.R., Daar, I.O., Krug, J.R., Maquat, L.E. (1985) Characterization of the functional gene and several processed pseudogenes in the human triosephosphate isomerase gene family. *Mol. Cell Biol.* 5:1694-1706.
34. Brown, P.O, Botstein, D. (1999) Exploring the new world of the genome with DNA microarrays. *Nature Genetics* 21:33-37.
35. Brunelleschi, S., Penengo, L., Santoro, M.M., Gaudino, G. (2002) Receptor tyrosine kinases as target for anti-cancer therapy. *Curr. Pharm. Des.* 8:1959-1972.
36. Brutlag, D.L., Dautricourt, J.P., Diaz, R., Fier, J., Moxon, B., Stamm, R. (1993). BLAZE: An implementation of the Smith-Waterman comparison algorithm on a massively parallel computer. *Computers and Chemistry* 17:203-207.

37. Carbonell, L.F., Hodge, M.R., Tomalski, M.D., Miller, L.K. (1988) Synthesis of a gene coding for an insect-specific scorpion neurotoxin and attempts to express it using baculovirus vectors. *Gene* 73:409-418.

38. Chakravarty, A. (1999) Population genetics - making sense out of sequence. *Nature Genetics* 21:56-60.

39. Chalut, C., Gallois, Y., Poterszman, A., Moncollin, V., Egly, J.M. (1995) Genomic structure of the human TATA-box-binding protein (TBP). *Gene* 161:277-282.

40. Chang, A.C., Nunberg, J.H., Kaufman, R.J., Erlich, H.A., Schimke, R.T., Cohen, S.N. (1978) Phenotypic expression in *E. coli* of a DNA sequence coding for mouse dihydrofolate reductase. *Nature* 275:617-624.

41. Chang, M.S., Chang, C.L., Huang, C.J., Yang, Y.C. (2000) p29, a novel GCIP-interacting protein, localizes in the nucleus. *Biochem. Biophys. Res. Commun.* 279:732-737.

42. Chen, F.W., Ioannou, Y.A. (1998) Ribosomal proteins in cell proliferation and apoptosis. *Int. Rev. Immunol.* 18:429-448.

43. Cheung, V.G., Morley, M., Aquilar, F., Massimi, A., Kucherlapati, R., Childs, G. (1999) Making and reading microarrays. *Nature Genetics* 21:15-19.

44. Christa, L., Simon, M.T., Flinois, J.P., Gebhardt, R., Brechot, C., Lasserre, C. (1994) Overexpression of glutamine synthetase in human primary liver cancer. *Gastroenterology* 106:1312-1320.

45. Clark, C.M., Karlawish, J.H. (2003) Alzheimer disease: current concepts and emerging diagnostic and therapeutic strategies. *Ann. Intern. Med.* 138:400-410.

46. Coffin, J.M., Hughes, S.H., Varmus, H.E. (1997) Retroviruses. Cold Spring Harbor Laboratory Press.

47. Cole, K.A., Krizman, D.B., Emmert-Buck, M.R. (1999) The genetics of cancer - a 3D model. *Nature Genetics* 21:38-41.

48. Collins, F.S. (1999) Microarrays and macroconsequences. *Nature Genetics* 21:2.

49. Comuzzie, A.G., Allison, D.B. (1998) The search for human obesity genes. *Science* 280:1374-1377.

50. Cormand, B., Montfort, M., Chabas, A., Vilageliu, L., Grinberg, D. (1997) Genetic fine localization of the beta-glucocerebrosidase (GBA) and prosaposin (PSAP) genes: implications for Gaucher disease. *Hum. Genet.* 100:75-79.

51. Cregg, J.M., Barringer, K.J., Hessler, A.Y., Madden, K.R. (1985) *Pichia* pastoris as a host system for transformations. *Mol. Cell. Biol.* 5:3376-3385.

52. Crooke, S.T. (1996) Progress in antisense therapeutics. *Med. Res. Rev.* 16:319-344.

53. Crouch, R.J. (1990) Ribonuclease H: from discovery to 3D structure. *New Biol.* 2:771-777.

54. Curcio, L.D., Bouffard, D.Y., Scanlon, K.J. (1997) Oligonucleotides as modulators of cancer gene expression. *Pharmacol. Ther.* 74:317-332.

55. Das, S., Kellermann, E., Hollenberg, C.P. (1984) Transformation of *Kluyveromyces fragilis*. *J. Bacteriol.* 158:1165-1167.

56. Davidow, L.S., Kaczmarek, F.S., DeZeeuw, J.R., Conlon, S.W., Lauth, M.R., Pereira, D.A., Franke, A.E. (1987) The Yarrowia lipolytica LEU2 gene. *Curr. Genet.* 11:377-383.

57. de Boer, H.A., Comstock, L.J., Vasser, M. (1993) The tac promoter: a functional hybrid derived from the trp and lac promoters. *Proc. Natl. Acad. Sci.* 80:21-25.

58. De Louvencourt, L., Fukuhara, H., Heslot, H., Wesolowski, M. (1983) Transformation of *Kluyveromyces lactis* by killer plasmid DNA. *J. Bacteriol.* 154:737-742.

59. Deasy, B.M., Huard, J. (2002) Gene therapy and tissue engineering based on muscle-derived stem cells. *Curr. Opin. Mol. Ther.* 4:382-389.

60. Deutscher, M.P., Simon, M.I., Abelson, J.N., eds. (1990) Guide to Protein Purification: Methods in Enzymology. (Methods in Enzymology Series, Vol 182). Academic Press.

61. Dieffenbach, C.W., Dveksler, G.S., eds. (1995) PCR Primer: A Laboratory Manual. Cold Spring Harbor Laboratory Press.

62. Dijkema, R., van der Meide, P.H., Pouwels, P.H., Caspers, M., Dubbeld, M., Schellekens, H. (1985) Cloning and expression of the chromosomal immune interferon gene of the rat. *EMBO J.* 4:761-767.

63. Doerfler, W., Bohm, P., eds. (1987) The Molecular Biology Of Baculoviruses. Springer-Verlag, Inc.

64. Doll, A., Grzeschik, K.H. (2001) Characterization of two novel genes, WBSCR20 and WBSCR22, deleted in Williams-Beuren syndrome. *Cytogenet. Cell Genet.* 95:20-27.

65. Doolittle, R.F., Abelson, J.N., Simon, M.I., eds. (1996) Computer Methods for Macromolecular Sequence Analysis. 1st ed. Academic Press.

66. Ducrest, A.L., Suzutorisz, H., Lingner, J., Nabholz, M. (2002) Regulation of the human telomerase reverse transcriptase gene. *Oncogene* 21:541-52.

67. Egilsson, V., Gudnason, V., Jonasdottir, A., Ingvarsson, S., Andresdottir, V. (1986) Catabolite repressive effects of 5-thio-D-glucose on *Saccharomyces cerevisiae*. *J. Gen. Microbiol.* 132:3309-3313.

68. Ehrhardt, G.R., Korherr, C., Wieler, J.S., Knaus, M., Schrader, J.W. (2001) A novel potential effector of M-Ras and p21 Ras negatively regulates p21 Ras-mediated gene induction and cell growth. *Oncogene* 20:188-197.

69. Espejo, A., Cote, J., Bednarek, A., Richard, S., Bedford, M.T. (2002) A protein-domain microarray identifies novel protein-protein interactions. *Biochem. J.* 367:697-702.

70. Everett, R.D., Meredith, M., Orr, A., Cross, A., Kathoria, M., Parkinson, J. (1997) A novel ubiquitin-specific protease is dynamically associated with the PML nuclear domain and binds to a herpesvirus regulatory protein. *EMBO J.* 16:1519-1530.

71. Fanning, A.S., Anderson, J.M. (1999) Protein modules as organizers of membrane structure. *Curr. Opin. Cell Biol.* 11:432-439.

72. Fisch, P., Forster, A., Sherrington, P.D., Dyer, M.J., Rabbitts, T.H. (1993) The chromosomal translocation t(X;14)(q28;q11) in T-cell pro-lymphocytic leukaemia breaks within one gene and activates another. *Oncogene* 8:3271-3276.

73. Fishman, P.S., Oyler, G.A. (2002) Significance of the parkin gene and protein in understanding Parkinson's disease. *Curr. Neurol. Neurosci. Rep.* 2:296-302.

74. Forgac, M. (1999) Structure and properties of the vacuolar (H+)-ATPases. *J. Biol. Chem.* 274:12,951-12,954.

75. Frank, I. (2002) Antivirals against HIV-1. *Clin. Lab. Med.* 22:741-757.

76. Frithz, G., Ericsson, P., Ronquist, G. (1976) Serum adenylate kinase activity in the early phase of acute myocardial infarction. *Ups J Med Sci.* 81:155-158.

77. Funakoshi, I., Kato, H., Horie, K., Yano, T., Hori, Y., Kobayashi, H., Inoue, T., Suzuki, H., Fukui, S., Tsukahara, M., et al. (1992) Molecular cloning of cDNAs for human fibroblast nucleotide pyrophosphatase. *Arch. Biochem. Biophys.* 295:180-187.

78. Gaillardin, C., Ribet, A.M. (1987) LEU2 directed expression of beta-galactosidase activity and phleomycin resistance in *Yarrowia lipolytica*. *Curr. Genet.* 11:369-375.

79. Gao, X., Nawaz, Z. (2002) Progesterone receptors - animal models and cell signaling in breast cancer: Role of steroid receptor coactivators and corepressors of progesterone receptors in breast cancer. *Breast Cancer Res.* 4:182-186.

80. Gao, Y., Melki, R., Walden, P.D., Lewis, S.A., Ampe, C., Rommelaere, H., Vandekerckhove, J., Cowan, N.J. (1994) A novel cochaperonin that modulates the ATPase activity of cytoplasmic chaperonin. *J. Cell Biol.* 125:989-996.

81. Geffen D.B., Man S. (2002) New drugs for the treatment of cancer, 1990-2001. *Isr. Med. Assoc. J.* 4:1124-31.

82. Ghofrani, H.A., Rose, F., Schermuly, R.T., Olschewski, H., Wiedemann, R., Kreckel, A., Weissmann, N., Ghofrani, S., Enke, B., Seeger, W., Grimminger, F. (2003) Oral sildenafil as long-term adjunct therapy to inhaled iloprost in severe pulmonary arterial hypertension. *J. Am. Coll. Cardiol.* 42:158-164.

83. Gillingham, A.K., Pfeifer, A.C., Munro, S. (2002) CASP, the alternatively spliced product of the gene encoding the CCAAT-displacement protein transcription factor, is a Golgi membrane protein related to giantin. *Mol. Biol. Cell* 13:3761-3774.

84. Gingras, M.C., Lapillonne, H., Margolin, J.F. (2002) TREM-1, MDL-1, and DAP12 expression is associated with a mature stage of myeloid development. *Mol. Immunol.* 38:817-824.

85. Girschick, H.J., Grammer, A.C., Nanki, T., Vazquez, E., Lipsky, P.E. (2002) Expression of recombination activating genes 1 and 2 in peripheral B cells of patients with systemic lupus erythematosus. *Arthritis. Rheum.* 46:1255-1263.

86. Gmeiner, W.H., Horita, D.A. (2001) Implications of SH3 domain structure and dynamics for protein regulation and drug design. *Cell Biochem. Biophys.* 35:127-140.

87. Goeddel, D.V., Heyneker, H.L., Hozumi, T., Arentzen, R., Itakura, K., Yansura, D.G., Ross, M.J., Mizzari, G., Crea, R., Seeburg, P.H. (1979) Direct expression in *E. coli* of a DNA sequence coding for human growth hormone. *Nature* 281:544-548.

88. Goldstein, L.S.B., Yang, Z. (2000) Microtubule-based transport systems in neurons: the roles of kinesins and dyneins. *Annu. Rev. Neurosci.* 23:39-71.

89. Golovkina, T.V., Chervonsky, A., Dudley, J.P., Ross, S.R. (1992) Transgenic moue mammary tumor virus superantigen expression prevents viral infection. *Cell* 69:637-645.

90. Gonnet, G.H., Cohen, M.A., Benner, S.A. (1992) Exhaustive matching of the entire protein sequence database. *Science* 256:1443-1445.

91. Gordan, J.D., Vonderheide, R.H. (2002) Universal tumor antigens as targets for immunotherapy. *Cytotherapy* 4:317-327.

92. Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I., Howard, B.H. (1982) The Rous sarcoma virus long terminal repeat is a strong promoter when introduced into a variety of eucaryotic cells by DNA-mediated transfection. *Proc. Natl. Acad. Sci.* 79:6777-6781.

93. Gray, T.A., Hernandez, L., Carey, A.H., Schaldach, M.A., Smithwick, M.J., Rus, K.M., Graves, J.A., Stewart, C.L., Nicholls, R.D. (2002) The ancient source of a distinct gene family encoding proteins featuring RING and C(3)H zinc-finger motifs with abundant expression in developing brain and nervous system. *Genomics*. 66:76-86.

94. Griffiths, A.J.F., Miller, J.H., Suzuki, D.T., Lewontin, R.C., Gelbart, W.M. (1999) Introduction to Genetic Analysis. 7th ed. W.H. Freeman.

95. Griffiths, M., Beaumont, N., Yao, S.Y., Sundaram, M., Boumah, C.E., Davies, A., Kwong, F.Y., Coe, I., Cass, C.E., Young, J.D., Baldwin, S.A. (1997) Cloning of a human nucleoside transporter implicated in the cellular uptake of adenosine and chemotherapeutic drugs. *Nat. Med.* 3:89-93.

96. Hacia, J.G. (1999) Resequencing and mutational analysis using oligonucleotide microarrays. *Nature Genetics* 21:42-47.

97. Hadano, S., Yanagisawa, Y., Skaug, J., Fichter, K., Nasir, J., Martindale, D., Koop, B.F., Scherer, S.W., Nicholson, D.W., Rouleau, G.A., Ikeda, J., Hayden, M.R. (2001) Cloning and characterization of three novel genes, ALS2CR1, ALS2CR2, and ALS2CR3, in the juvenile amyotrophic lateral

sclerosis (ALS2) critical region at chromosome 2q33-q34: candidate genes for ALS2. *Genomics* 71:200-213.

98. Hall, M., Mickey, D.D., Wenger, A.S., Silverman, L.M. (1985) Adenylate kinase: an oncodevelopmental marker in an animal model for human prostatic cancer. *Clin. Chem.* 31:1689-1691.

99. Ham, R.G., McKeehan, W.L. (1979) Media and growth requirements. *Methods Enzymol.* 58:44-93.

100. Hanada, T., Lin, L., Tibaldi, E.V., Reinherz, E.L., Chishti, A.H. (2000) GAKIN, a novel kinesin-like protein associates with the human homologue of the Drosophila discs large tumor suppressor in T lymphocytes. *J. Biol. Chem.* 275:28,774-28,784.

101. Hartmann, G., Endres, S., eds. (1999) Manual of Antisense Methodology (Perspectives in Antisense Science). 1st ed. Kluwer Law International.

102. Hawes, J.W., Jaskiewicz, J., Shimomura, Y., Huang, B., Bunting, J., Harper, E.T., Harris, R.A. (1996) Primary structure and tissue-specific expression of human beta-hydroxyisobutyryl-coenzyme A hydrolase. *J. Biol. Chem.* 271:26,430-26,434.

103. Heath, J.K., White, S.J., Johnstone, C.N., Catimel, B., Simpson, R.J., Moritz, R.L., Tu, G.F., Ji, H., Whitehead, R.H., Groenen, .L.C., Scott, A.M., Ritter, G., Cohen, L., Welt, S., Old, L.J., Nice, E.C., Burgess, A.W. (1997) The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. *Proc. Natl. Acad. Sci.* 94:469-474.

104. Henningson, C.T. Jr., Stanislaus, M.A., Gewirtz, A.M. (2003) Embryonic and adult stem cell therapy. *J. Allergy Clin. Immunol.* 111:S745-S753.

105. Hinnen, A., Hicks, J.B., Fink, G.R. (1978) Transformation of yeast. *Proc. Natl. Acad. Sci.* 75:1929-1933.

106. Hirsch, D.S., Pirone, D.M., Burbelo, P.D. (2001) A new family of Cdc42 effector proteins, CEPs, function in fibroblast and epithelial cell shape changes. *J. Biol. Chem.* 276:875-883.

107. Ho, L.W., Carmichael, J., Swartz, J., Wyttenbach, A., Rankin, J., Rubinsztein, D.C. (2001) The molecular biology of Huntington's disease. *Psychol. Med.* 31:3-14.

108. Hollis, G.F., Evans, R.J., Stafford-Hollis, J.M., Korsmeyer, S.J., McKearn, J.P. (1989) Immunoglobulin lambda light-chain-related genes 14.1 and 16.1 are expressed in pre-B cells and may encode the human immunoglobulin omega light-chain protein. *Proc. Natl. Acad. Sci.* 86:5552-5556.

109. Hoozemans, J.J., Veerhuis, R., Rozemuller, A.J., Eikelenboom, P. (2002) The pathological cascade of Alzheimer's disease: the role of inflammation and its therapeutic implications. *Drugs Today (Barc)* 38:429-443.

110. Houseman, B.T., Huh, J.H., Kron, S.J., Mrksich, M. (2002) Peptide chips for the quantitative evaluation of protein kinase activity. *Nature Biotechnol.* 20:270-274.

111. Huynh, D.P., Yang, H.T., Vakharia, H., Nguyen, D., Pulst, S.M. (2003) Expansion of the polyQ repeat in ataxin-2 alters its Golgi localization, disrupts the Golgi complex and causes cell death. *Hum. Mol. Genet.* 12:1485-1496.

112. Ikeda, A., Nishina, P.M., Naggert, J.K. (2002) The tubby-like proteins, a family with roles in neuronal development and function. *J. Cell Sci.* 115(Pt 1):9-14.

113. Ito, H., Fukuda, Y., Murata, K., Kimura, A. (1978) Transformation of intact yeast cells treated with alkali cations. *J. Bacteriol.* 153:163-168.

114. Janeway, C.A., Travers, P. Walport, M. Shlomchik, M. (2001) Immunobiology. 5th ed. Garland Publishing.

115. Jeffery, P., Zhu, J. (2002) Mucin-producing elements and inflammatory cells. *Novartis Found. Symp.* 248:51-75, 277-82.

116. Jimbo, T., Kawasaki, Y., Koyama, R., Sato, R., Takada, S., Haraguchi, K., Akiyama, T. (2002) Identification of a link between the tumour suppressor APC and the kinesin superfamily. *Nat. Cell Biol.* 4:323-327.

117. Joberty, G., Perlungher, R.R., Macara, I.G. (1999) The Borgs, a new family of Cdc42 and TC10 GTPase-interacting proteins. *Mol. Cell Biol.* 19:6585-6597.

118. Johns, T.G., Bernard, C.C. (1997) Binding of complement component Clq to myelin oligodendrocyte glycoprotein: a novel mechanism for regulating CNS inflammation. *Mol. Immunol.* 34:33-38.

119. Jolliffe, C.N., Harvey, K.F., Haines, B.P., Parasivam, G., Kumar, S. (2000) Identification of multiple proteins expressed in murine embryos as binding partners for the WW domains of the ubiquitin-protein ligase Nedd4. *Biochem. J.* 351:557-565.

120. Jones, P., ed. (1998a) Vectors: Cloning Applications: Essential Techniques, John Wiley & Son, Ltd.

121. Jones, P., ed. (1998b) Vectors: Expression Systems: Essential Techniques, John Wiley & Son, Ltd.

122. Jurcic, J.G., Cathcart, K., Pinilla-Ibarz, J., Scheinberg, D.A. (2000) Advances in immunotherapy of hematlogic malignancies: cellular and humoral approaches. *Curr. Opin. Hematol.* 7:247-254.

123. Jury, J.A., Perry, A.C., Hall, L. (1999) Identification, sequence analysis and expression of transcripts encoding a putative metalloproteinase, eMDC II, in human and macaque epididymis. *Mol. Hum. Reprod.* 5:1127-1134.

124. Kamitani, T., Nguyen, H.P., Yeh, E.T. (1997) Preferential modification of nuclear proteins by a novel ubiquitin-like molecule. *J. Biol. Chem.* 272:14,001-14,004.

125. Kantoff, P.W., Halabi, S., Farmer, D.A., Hayes, D.F., Vogelzang, N.A., Small, E.J. (2001) Prognostic significance of reverse transcriptase polymerase chain reaction for prostate-specific antigen in men with hormone-refractory prostate cancer. *J. Clin. Oncol.* 9:3025-3028.

126. Kao, P.N., Chen, L., Brock, G., Ng, J., Kenny, J., Smith, A.J., Corthesy, B. (1994) Cloning and expression of cyclosporin A- and FK506-sensitive nuclear factor of activated T-cells: NF45 and NF90. *J. Biol. Chem.* 269:20,691-20,699.

127. Karanazanashvili, G., Abrahamsson, P. (2003) Prostate specific antigen and human glandular kallikrein 2 in early detection of prostate cancer. *J. Urol.* 169:445-457.

128. Kari, C., Chan, T.O., Rocha de Quadros, M., Rodeck, U. (2003) Targeting the epidermal growth factor receptor in cancer: apoptosis takes center stage. *Cancer Res.* 63:1-5.

129. Kelly, J.M., Hynes, M.J. (1985) Transformation of *Aspergillus niger* by the mdS gene of *Aspergillus nidulans*. *EMBO J.* 4:475-479.

130. Kenmochi, N., Kawaguchi, T., Rozen, S., Davis, E., Goodman, N., Hudson, T.J., Tanaka, T., Page, D.C. (1998) A map of 75 human ribosomal protein genes. *Genome Res.* 8:509-523.

131. Kirkpatrick, K.L., Mokbel, K. (2001) The significance of human telomerase reverse transcriptase (hTERT) in cancer. *Eur. J. Surg. Oncol.* 27:754-760.

132. Kirsch, K.H., Georgescu, M.M., Ishimaru, S., Hanafusa, H. (1999) CMS: an adapter molecule involved in cytoskeletal rearrangements. *Proc. Natl. Acad. Sci.* 96:6211-6216.

133. Kiryu-Seo, S., Sasaki, M., Yokohama, H., Nakagomi, S., Hirayama, T., Aoki, S., Wada, K., Kiyama, H. (2000) Damage-induced neuronal endopeptidase (DINE) is a unique metallopeptidase expressed in response to neuronal damage and activates superoxide scavengers. *Proc. Natl. Acad. Sci.* 97:4345-4350.

134. Klarman, G.J., Hawkins, M.E., Le Grice, S.F. (2002) Uncovering the complexities of retroviral ribonuclese H reveals its potential as a therapeutic target. *AIDS Rev.* 4:183-194.

135. Kobayashi, M., Takezawa, S., Hara, K., Yu, R.T., Umesono, Y., Agata, K., Taniwaki, M., Yasuda, K., Umesono, K. (1999) Identification of a photoreceptor cell-specific nuclear receptor. *Proc. Natl. Acad. Sci.* 96:4814-4819.

136. Korner, C., Knauer, R., Stephani, U., Marquardt, T., Lehle, L., von Figura, K. (1999) Carbohydrate deficient glycoprotein syndrome type IV: deficiency of dolichyl-P-Man:Man(5)GlcNAc(2)-PP-dolichyl mannosyltransferase. *EMBO J.* 18:6816-6822.

137. Kothapalli, R., Buyuksal, I., Wu, S.Q., Chegini, N., Tabibzadeh, S. (1997) Detection of ebaf, a novel human gene of the transforming growth factor beta superfamily association of gene expression with endometrial bleeding. *J. Clin. Invest.* 99:2342-2350.

138. Kovalenko, O.V., Golub, E.I., Bray-Ward, P., Ward, D.C., Radding, C.M. (1997) A novel nucleic acid-binding protein that interacts with human rad51 recombinase. *Nucleic Acids Res.* 25:4946-4953.

139. Kratzschmar, J., Lum, L., Blobel, C.P. (1996) Metargidin, a membrane-anchored metalloprotease-disintegrin protein with an RGD integrin binding sequence. *J. Biol. Chem.* 271:4593-4596.

140. Ku, D.H., Kagan, J., Chen, S.T., Chang, C.D., Baserga, R., Wurzel, J. (1990) The human fibroblast adenine nucleotide translocator gene. Molecular cloning and sequence. *J. Biol. Chem.* 265:16,060-16,063.

141. Kuisle, O., Quiñoá, E., Rigura, R. (1999) Solid phase synthesis of depsides and depsipeptides. *Tetrahedron Lett.* 40:1203-1206.

142. Kunze, G. et al., (1985) Transformation of the industrially important yeasts *Candida* maltosa and *Pichia* guilliermondii. *J. Basic Microbiol.* 25:141-144.

143. Kurtz, M.B., Cortelyou, M.W., Kirsch, D.R. (1986) Integrative transformation of Candida albicans, using a cloned *Candida* ADE2 gene. *Mol. Cell. Biol.* 6:142-149.

144. Kyo, S., Takakura, M., Inoue, M. (2000) Telomerase activity in cancer as a diagnostic and therapeutic target. *Histol. Histopathol.* 15:813-824.

145. Lander, E.S. (1999) Array of hope. *Nature Genetics* 21:3-4.

146. Lander, E.S., Linton, L..M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R.., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S.., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendl, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P.,

Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Smith, D.R., Doucette-Stamm, L., Rubenfield, M., Weinstock, K., Lee, H.M., Dubois, J., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Abola, A.P., Proctor, M.J., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W.R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Johnson, L.S., Jones, T.A., Kasif, S., Kaspryzk, A., Kennedy, S., Kent, W.J., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F., Stupka, E., Szustakowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.P., Yeh, R.F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Patrinos, A., Morgan, M.J., Szustakowki, J., de Jong, P., Catanese, J.J., Osoegawa, K., Shizuya, H., Choi, S., Chen, Y.J.; International Human Genome Sequencing Consortium. (2001) Initial sequencing and analysis of the human genome *Nature* 409:860-921.

147. Lasham, A., Moloney, S., Hale, T., Homer, C., Zhang, Y.F., Murison, J.G., Braithwaite, A.W., Watson, J. (2003) The Y-box binding protein YB1:

A potential negative regulator of the p53 tumor suppressor. *J. Biol. Chem.* Epub ahead of print, June 30, 2003.

148. Lashkari, A., Smith, A.K., Graham, J.M. Jr. (1999) Williams-Beuren syndrome: an update and review for the primary physician. *Clin. Pediatr.* 38:189-208.

149. Lavedan, C. (1998) The synuclein family. *Genome Res.* 8:871-880.

150. Lebacq-Verheyden, A.M., Kasprzyk, P.G., Raum, M.G., Van Wyke Coelingh, K., Lebacq, J.A., Battey, J.F. (1988) Posttranslational processing of endogenous and of baculovirus-expressed human gastrin-releasing peptide precursor. *Mol. Cell. Biol.* 8:3129-3135.

151. Lees-Miller, S.P., Anderson, C.W. (1989) Two human 90-kDa heat-shock proteins are phosphorylated in vivo at conserved serines that are phosphorylated *in vitro* by casein kinase II. *J. Biol. Chem.* 264:2431-2437.

152. Lerch, M.M., Gorelick, F.S. (2000) Early trypsinogen activation in acute pancreatitis. *Med. Clin. North Amer.* 84:549-563.

153. Lerner, R.A. (1982) Tapping the immunological repertoire to produce antibodies of predetermined specificity. *Nature* 299:592-596.

154. Li, E., Bestagno, M., Burrone, O. (1996) Molecular cloning and characterization of a transmembrane surface antigen in human cells. *Eur. J. Biochem.* 238:631-638.

155. Lim, D., Orlova, M., Goff, S.P. (Aug. 2002) Mutations of the RNase H C helix of the Moloney murine leukemia virus reverse transcriptase reveal defects in polypurine tract recognition. *J. Virol.* 76:8360-8373.

156. Lin, B., Rommens, J.M., Graham, R.K., Kalchman, M., MacDonald, H., Nasir, J., Delaney, A., Goldberg, Y.P., Hayden, M.R. (1993) Differential 3'polyadenylation of the Huntington disease gene results in two mRNA species with variable tissue expression. *Hum. Mol. Genet.* 2:1541-1545.

157. Lin, W.J., Gary, J.D., Yang, M.C., Clarke, S., Herschman, H.R. (1996) The mammalian immediate-early TIS21 protein and the leukemia-associated BTG1 protein interact with a protein-arginine N-methyltransferase. *J. Biol. Chem.* 271:15,034-15,044.

158. Lin, X., Sikkink, R.A., Rusnak, F., Barber, D.L. (1999) Inhibition of calcineurin phosphatase activity by a calcineurin B homologous protein. *J. Biol. Chem.* 274:36,125-36,131.

159. Linnenbach, A.J., Seng, B.A., Wu, S., Robbins, S., Scollon, M., Pyrc, J.J., Druck, T., Huebner, K. (1993) Retroposition in a family of carcinoma-associated antigen genes. *Mol. Cell Biol.* 13:1507-1515.

160. Linstedt, A.D., Hauri, H.P. (1993) Giantin, a novel conserved Golgi membrane protein containing a cytoplasmic domain of at least 350 kDa. *Mol. Biol. Cell* 4:679-693.

161. Lipshutz, R.J., Fodor, S.P.A., Gingeras, T.R., Lockhart, D.J. (1999) High density synthetic oligonucleotide arrays. *Nature Genetics* 21:20-24.

162. Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D., Darness, J. (1999) Molecular Cell Biology. 4th ed. W H Freeman & Co.

163. Loeffen, J.L., Triepels, R.H., van den Heuvel, L.P., Schuelke, M., Buskens, C.A., Smeets, R.J., Trijbels, J.M., Smeitink, J.A. (1998) cDNA of eight nuclear encoded subunits of NADH:ubiquinone oxidoreductase: human complex I cDNA characterization completed. *Biochem. Biophys. Res. Commun.* 253:415-422.

164. Los, M., Burek, C.J., Stroh, C., Benedyk, K., Hug, H., Mackiewicz. (2003) Anticancer drugs of tomorrow: apoptotic pathways as targets for drug design. *Drug Discov. Today* 15:67-77.

165. Lovering R, Trowsdale J. (1991) A gene encoding 22 highly related zinc fingers is expressed in lymphoid cell lines. *Nucleic Acids Res.* 19:2921-2928.

166. Luckow, V., Summers, M. (1988) Trends in the development of baculovirus expression vectors. *Bio/Technology* 6:47-55.

167. MacBeath, G., Schreiber. S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289:1760-1763.

168. Machesky, L.M., Reeves, E., Wientjes, F., Mattheyse, F.J., Grogan, A., Totty, N.F., Burlingame, A.L., Hsuan, J.J., Segal, A.W. (1999) Mammalian actin-related protein 2/3 complex localizes to regions of lamellipodial protrusion and is composed of evolutionarily conserved proteins. *Biochem. J.* 328:105-112.

169. Mackay, A., Jones, C., Dexter, T., Silva, R.L., Bulmer, K., Jones, A., Simpson, P., Harris, R.A., Jat, P.S., Neville, A.M., Reis, L.F., Lakhani, S.R., O'Hare, M.J. (2003) cDNA microarray analysis of genes associated with

ERBB2 (HER2/neu) overexpression in human mammary luminal epithelial cells. *Oncogene* 22:2680-2688.

170. Maeda, S., Kawai, T., Obinata, M., Fujiwara, H., Horiuchi, T., Saeki, Y., Sato, Y., Furusawa, M. (1985) Production of human alpha-interferon in silkworm using a baculovirus vector. *Nature* 315:592-594.

171. Mahajan, M.A., Murray, A., Samuels, H.H. (2002) NRC-interacting factor 1 is a novel cotransducer that interacts with and regulates the activity of the nuclear hormone receptor coactivator NRC. *Mol. Cell Biol.* 22:6883-6894.

172. Mahimkar, R.M., Baricos, W.H., Visaya, O., Pollock, A.S., Lovett, D.H. (2000) Identification, cellular distribution and potential function of the metalloprotease-disintegrin MDC9 in the kidney. *J. Am. Soc. Nephrol.*, 11:595-603.

173. Mahnensmith, R.L., Aronson, P.S. (1985) Interrelationships among quinidine, amiloride, and lithium as inhibitors of the renal Na+-H+ exchanger. *J. Biol. Chem.* 260:12,586-12,592.

174. Manning, G., Whyte, D.B., Martinez, R., Hunter, T., Sudarsanam, S. (2002) The protein kinase complement of the human genome. *Science* 298:1912-1934.

175. Martel-Pelletier, J., Welsch, D.J., and Pelleteir, J.P. (2001) Metalloproteases and inhibitors in arthritic diseases. *Best Pract. Res. Clin. Rheumatol.* 15:805-829.

176. Martin, B.M., Tsuji, S., LaMarca, M.E., Maysak, K., Eliason, W., Ginns, E.I. (1988) Glycosylation and processing of high levels of active human glucocerebrosidase in invertebrate cells using a baculovirus expression vector. *DNA* 7:99-106.

177. Massari, M.E., Rivera, R.R., Voland, J.R., Quong, M.W., Breit, T.M., van Dongen, J.J,, de Smit, O., Murre, C. (1998) Characterization of ABF-1, a novel basic helix-loop-helix transcription factor expressed in activated B lymphocytes. *Mol. Cell Biol.* 18:3130-3139.

178. Matz, M.V., Fradkov, A.F., Labas, Y.A., Savitsky, A.P., Zaraisky, A.G., Markelov, M.L., Lukyanov, S.A. (1999) Fluorescent proteins from nonbioluminescent *Anthozoa* species. *Nat. Biotechnol.* 17:969-973.

179. Mayer, B.J. (2001) SH3 domains: complexity in moderation. *J. Cell Sci.* 114:1253-1263.

180. Mayer, T.U., Kapoor, T.M., Haggarty, S.J., King, R.W., Schreiber, S.L., Mitchison, T.J. (1999) Small molecule inhibitor of mitotic spindle bipolarity identified in a phenotype-based screen. *Science* 286:971-974.

181. McKusick, V.A.. (2003) OMIM: Online Mendelian Inheritance in Man http:www.ncbi.nlm.nih.gov, #104300.

182. McPherson, M.J., Møller, S.G., Benyon, R., Howe, C. (2000) PCR Basics: From Background to Bench. Springer Verlag.

183. Merla, G., Ucla, C., Guipponi, M., Reymond, A. (2002) Identification of additional transcripts in the Williams-Beuren syndrome critical region. *Hum. Genet.* 110:429-438.

184. Miki, H., Setou, M., Kaneshiro, K., Hirokawa, N. (2001) All kinesin superfamily protein, KIF, genes in mouse and human. *Proc. Natl. Acad. Sci.* 98:7004-7011.

185. Milam, A.H., Rose, L., Cideciyan, A.V., Barakat, M.R., Tang, W.X., Gupta, N., Aleman, T.S., Wright, A.F., Stone, E.M., Sheffield, V.C., Jacobson, S.G. (2002) The nuclear receptor NR2E3 plays a role in human retinal photoreceptor differentiation and degeneration. *Proc. Natl. Acad. Sci.* 99:473-478.

186. Mitch, W.E., Goldberg, A.L. (1996) Mechanisms of muscle wasting. The role of the ubiquitin-proteasome pathway. *N. Engl. J. Med.* 335:1897-1905.

187. Miyajima A. (2002) Functional analysis of yeast homologue gene associated with human DNA helicase causative syndromes. *Kokuritsu Iyakuhin Shokuhin Eisei Kenkyusho Hokoku* 120:53-74.

188. Miyajima, A., Schreurs, J., Otsu, K., Kondo, A., Arai, K., Maeda, S. (1987) Use of the silkworm, *Bombyx mori*, and an insect baculovirus vector for high-level expression and secretion of biologically active mouse interleukin-3. *Gene* 58:273-281.

189. Monfardini, C., Schiavon, O., Caliceti, P., Morpurgo, M., Harris, J.M., Veronese, F.M. (1995) A branched monomethoxypoly(ethylene glycol) for protein modification. *Bioconjugate Chem.* 6:62-69.

190. Mori, N. (1997) Neuronal growth-associated proteins in neural plasticity and brain aging. *Nihon Shinkei Seishin Yakurigaku Zasshi* 17:159-167.

191. Myers, E.W., Miller, W. (1988) Optimal alignments in linear space. *Comput. Appl. Biosci.* 4:11-7.

192. Nagata, K., Kawase, H., Handa, H., Yano, K., Yamasaki, M., Ishimi, Y., Okuda, A., Kikuchi, A., Matsumoto, K. (1995) Replication factor encoded by a putative oncogene, set, associated with myeloid leukemogenesis. *Proc. Natl. Acad. Sci.* 92:4279-4283.

193. Naora, H. (1999) Involvement of ribosomal proteins in regulating cell growth and apoptosis: translational modulation or recruitment for extraribosomal activity? *Immunol. Cell Biol.* 77:197-205.

194. Needleman, S.B., Wunch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* 48:443-453.

195. Nelson, N., Harvey, W.R. (1999) Vacuolar and plasma membrane proton-adenosine triphosphatases. *Physiol. Rev.* 79:361-385.

196. Nishiyama, H., Higashitsuji, H., Yokoi, H., Itoh, K., Danno, S., Matsuda, T., Fujita, J. (1997) Cloning and characterization of human CIRP (cold-inducible RNA-binding protein) cDNA and chromosomal assignment of the gene. *Gene* 204:115-120.

197. Noma, T., Fujisawa, K., Yamashiro, Y., Shinohara, M., Nakazawa, A., Gondo, T., Ishihara, T., Yoshinobu, K. (2001) Structure and expression of human mitochondrial adenylate kinase targeted to the mitochondrial matrix. *Biochem. J.* 358:225-232.

198. Notredame, C., Higgins, D., Heringa, J. (2000) T-Coffee: A novel method for multiple sequence alignments. *J. Molec. Biol.* 302:205-217.

199. Okazaki, Y., Furuno, M., Kasukawa. T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Bradt, D., Brusic, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierski, R.M., King, B.L., Konagaya, A., Kurochkin, IV, Lee, Y., Lenhard, B., Lyons,

P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E., Hayashizaki, Y.; FANTOM Consortium; RIKEN Genome Exploration Research Group Phase I & II Team. (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420:563-573.

200. Oksenberg, J.R., Barcellos, L.F., Hauser, S.L. (1999) Genetic aspects of multiple sclerosis. *Semin. Neurol.* 19:281-288.

201. Oliver, C.J., Shenolikar, S. (1998) Physiologic importance of protein phosphatase inhibitors. *Frontiers in Bioscience* 3:961-972.

202. O'Neil, N.J., Martin, R.L., Tomlinson, M.L., Jones, M.R., Coulson, A., Kuwabara, P.E. (2001) RNA-mediated interference as a tool for identifying drug targets. *Am. J. Pharmacogenomics* 1:45-53.

203. Page, D.C., Silber, S., Brown, L.G. (1999) Men with infertility caused by AZFc deletion can produce sons by intracytoplasmic sperm injection, but are likely to transmit the deletion and infertility. *Hum. Reprod.* 14:1722-1726.

204. Pan, C.X., Koeneman, K.S. (1999) A novel tumor-specific gene therapy for bladder cancer. *Med. Hypothesis* 53:130-135.

205. Pang, T., Wakabayashi, S., Shigekawa, M. (2001) Calcineurin homologous protein as an essential cofactor for Na+/H+ exchangers. *J. Biol. Chem* 276:17,367-17,372.

206. Pang, T., Wakabayashi, S., Shigekawa, M. (2002) Expression of calcineurin B homologous protein 2 protects serum deprivation-induced cell

death by serum-independent activation of Na+/H+ exchanger. *J. Biol. Chem.* 277:43,771-43,777.

207. Papagerakis, S., Shabana, A.H., Depondt, J., Gehanno, P., Forest, N. (2003) Immunohistochemical localization of plakophilins (PKP1, PKP2, PKP3, and p0071) in primary oropharyngeal tumors: correlation with clinical parameters. *Hum. Pathol.* 34:565-572.

208. Pearson, W.R. (2000) Flexible sequence similarity searching with the FASTA3 program package. *Methods Mol. Biol.* 132:185-219.

209. Peattie, D.A., Harding, M.W., Fleming, M.A., DeCenzo, M.T., Lippke, J.A., Livingston, D.J., Benasutti, M. (1992) Expression and characterization of human FKBP52, an immunophilin that associates with the 90-kDa heat-shock protein and is a component of steroid receptor complexes. *Proc. Natl. Acad. Sci.* 89:10,974-10,978.

210. Peelle, B., Gururaja, T.L., Payan, D.G., Anderson, D.C. (2001) Characterization and use of green fluorescent proteins from *Renilla mulleri* and *Ptilosarcus guernyi* for the human cell display of functional peptides. *J. Protein Chem.* 20:507-519.

211. Pepin, K., Momose, F., Ishida, N., Nagata, K. (2001) Molecular cloning of horse Hsp90 cDNA and its comparative analysis with other vertebrate Hsp90 sequences. *J. Vet. Med. Sci.* 63:115-124.

212. Perez Calvo, J.I., Inigo Gil, P., Giraldo Castellano, P., Torralba Cabeza, M.A., Civeira, F., Lario Garcia, S., Pocovi, M., Lara Garcia, S. (2000) Transforming growth factor beta (TGF-beta) in Gaucher's disease. Preliminary results in a group of patients and their carrier and non-carrier relatives *Med. Clin. (Barc)* 115:601-604.

213. Perron, H., Garson, J.A., Bedin, F., Beseme, F., Paranhos-Baccala, G., Komurian-Pradel, F., Mallet, F., Tuke, P.W., Voisset, C., Blond, J.L., Lalande, B., Seigneurin, J.M., Mandrand, B., The Collaborative Research Group on Multiple Sclerosis (1997) Molecular identification of a novel retrovirus repeatedly isolated from patients with multiple sclerosis. *Proc. Natl. Acad. Sci.* 94:7583-7588.

214. Perry, A.C., Jones, R., Hall, L. (1995) Analysis of transcripts encoding novel members of the mammalian metalloprotease-like, disintegrin-

like, cysteine-rich (MDC) protein family and their expression in reproductive and non-reproductive monkey tissues. *Biochem. J.* 312(Pt 1):239-244.

215. Pfutzer, R.H., Whitcomb, D.C. (2001) SPINK1 mutations are associated with multiple phenotypes. *Pancreatology* 1:457-460.

216. Phillips, M.I., ed. (1999a) Antisense Technology, Part A. Methods in Enzymology Vol. 313. Academic Press, Inc.

217. Phillips, M.I., ed. (1999b) Antisense Technology, Part B. Methods in Enzymology Vol. 314. Academic Press, Inc.

218. Pisegna, J.R., Wank, S.A. (1996) Cloning and characterization of the signal transduction of four splice variants of the human pituitary adenylate cyclase activating polypeptide receptor. Evidence for dual coupling to adenylate cyclase and phospholipase C. *J. Biol. Chem.* 271:17,267-17,274.

219. Price, N.T., Hall, L., Proud, C.G. (1993) Cloning of cDNA for the beta-subunit of rabbit translation initiation factor-2 using PCR. *Biochim. Biophys. Acta* 1216:170-172.

220. Qin, J., Li., L. (2003) Molecular anatomy of the DNA damage and replication checkpoints. *Radiat. Res.* 159:139-148.

221. Racevskis, J., Dill, A., Stockert, R., Fineberg, S.A. (1996) Cloning of a novel nucleolar guanosine 5'-triphosphate binding protein autoantigen from a breast tumor. *Cell. Growth Differ.* 7:271-280.

222. Ramalho-Santos, M. (2002) "Stemness" *Science* 298:597-600.

223. Rebbe, N.F., Ware, J., Bertina, R.M., Modrich, P., Stafford, D.W. (1987) Nucleotide sequence of a cDNA for a member of the human 90-kDa heat-shock protein family. *Gene* 53:235-245.

224. Rechid, R., Vingron, M., Argos, P. (1989) A new interactive protein sequence alignment program and comparison of its results with widely used algorithms. *Comput. Appl. Biosci.* 5:107-113.

225. Rehli, M., Krause, S.W., Kreutz, M., Andreesen, R. (1995) Carboxypeptidase M is identical to the MAX.1 antigen and its expression is associated with monocyte to macrophage differentiation. *J. Biol. Chem.* 270:15644-15649.

226. Ribardo, D.A., Peterson, J.W., Chopra, A.K. (2002) Phospholipase A2-activating protein--an important regulatory molecule in modulating

cyclooxygenase-2 and tumor necrosis factor production during inflammation. *Indian J. Exp. Biol.* 40:129-138.

227. Ritter, R.C., Brenner, L.A., Tamura, C.S. (1994) Endogenous CCK and the peripheral neural substrates of intestinal satiety. *Ann. N. Y. Acad. Sci.* 713:255-267.

228. Robertson, H.M. (1996) Members of the pogo superfamily of DNA-mediated transposons in the human genome. *Mol. Gen. Genet.* 252:761-766.

229. Robertson, H.M., Zumpano, K.L. (1997) Molecular evolution of an ancient mariner transposon, Hsmar1, in the human genome. *Gene* 205:203-217.

230. Roepman, R., Bernoud-Hubac, N., Schick, D.E., Maugeri, A., Berger, W., Ropers, H.H., Cremers, F.P., Ferreira, P.A. (2000) The retinitis pigmentosa GTPase regulator (RPGR) interacts with novel transport-like proteins in the outer segments of rod photoreceptors. *Hum. Mol. Genet.* 9:2095-2105.

231. Roessler, B.J., Nosal, J.M., Smith, P.R., Heidler, S.A., Palella, T.D., Switzer, R.L., Becker, M.A. (1993) Human X-linked phosphoribosylpyrophosphate synthetase superactivity is associated with distinct point mutations in the PRPS1 gene. *J. Biol. Chem.* 268:26476-26481.

232. Roggenkamp, R., Janowicz, Z., Stanikowski, B., Hollenberg, C.P. (1984) Biosynthesis and regulation of the peroxisomal methanol oxidase from the methylotrophic yeast *Hansenula polymorpha*. *Mol. Gen. Genet.* 194:489-493.

233. Rosen, R.C., McKenna, K.E. (2002) PDE-5 inhibition and sexual response: pharmacological mechanisms and clinical outcomes. *Ann. Rev. Sex Res.* 13:36-88.

234. Rosato, R.R., Grant, S. (2003) Histone deacetylase inhibitors in cancer therapy. *Cancer Biol. Ther.* 2:30-37.

235. Rowland, J.M. (2002) Molecular genetic diagnosis of pediatric cancer: current and emerging methods. *Pediatr. Clin. North Am.* 49:1415-1435.

236. Saha, S., Bardelli, A., Buckhaults, P., Velculescu, V.E., Rago, C., St Croix, B., Romans, K.E., Choti, M.A., Lengauer, C., Kinzler, K.W.,

Vogelstein, B. (2001) A phosphatase associated with metastasis of colorectal cancer. *Science* 294:1343-1346.

237. Saiki, R.K, Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., Erlich, H.A. (1988) Primer-directed enzymatic amplification of DNA with amplification of DNA with a thermostable DNA polymerase. *Science* 239:487-491.

238. Sambrook, J., Russell, D.W., Sambrook, J. (1989) Molecular Cloning, A Laboratory Manual. 2nd ed. Cold Spring Harbor Laboratory Press.

239. Sanchez, E.R., Faber, L.E., Henzel, W.J., Pratt, W.B. (1990) The 56-59-kilodalton protein identified in untransformed steroid receptor complexes is a unique protein that exists in cytosol in a complex with both the 70- and 90-kilodalton heat-shock proteins. *Biochemistry* 29:5145-5152.

240. Schaeferling, M., Schiller, S., Paul, H., Kruschina, M., Pavlickova, M., Meerkamp, M., Giammasi, C., Kambhampati, D. (2002) Application of self-assembly techniques in the design of biocompatible protein microarray surfaces. *Electrophoresis* 23:3097-3105.

241. Schaffer, J.E., Lodish, H.F. (1994) Expression cloning and characterization of a novel adipocyte long chain fatty acid transport protein. *Cell* 79:393-395.

242. Schena, M., ed. (1999) DNA Microarrays: A Practical Approach. Oxford Univ. Press.

243. Schena, M., ed. (2000) Microarray Biochip Technology. 1st ed. Eaton Publishing Co.

244. Schlesinger, D.H. (1988a) MacRomolecular Sequencing and Synthesis: Selected Methods and Applications. Wiley-Liss.

245. Schlesinger, D.H., ed. (1988b) Current Methods in Sequence Comparison and Analysis, Macromolecule Sequencing and Synthesis, Selected Methods and Applications, pp. 127-149, Alan R. Liss, Inc.

246. Schonthal, A.H. (2001) Role of serine/threonine protein phosphatase 2A in cancer. *Cancer Lett.* 170:1-13.

247. Seelig, H.P., Schranz, P., Schroter, H., Wiemann, C., Renz, M. (1994) Macrogolgin--a new 376 kD Golgi complex outer membrane protein as target of antibodies in patients with rheumatic diseases and HIV infections. *J. Autoimmun.* 7:67-91.

248. Selkoe, D.J. (2001) Presenilin, Notch, and the genesis and treatment of Alzheimer's disease. *Proc. Natl. Acad. Sci.* 98:11,039-11,041.

249. Setlow, J., Hollaender, A., eds. (1986) Genetic Engineering: Principles and Methods. Plenum Pub. Corp.

250. Shamay, M., Barak, O., Doitsh, G., Ben-Dor, I., Shaul, Y. (2002) Hepatitis B virus pX interacts with HBXAP, a PHD finger protein to coactivate transcription. *J. Biol. Chem.* 277:9982-9988.

251. Shao, H., Andres, D.A. (2000) A novel RalGEF-like protein, RGL3, as a candidate effector for rit and Ras. *J. Biol. Chem.* 275:26,914-26,924.

252. Sheppard, P., Kindsvogel, W., Xu, W., Henderson, K., Schlutsmeyer, S., Whitmore, T.E., Kuestner, R., Garrigues, U., Birks, C., Roraback, J., Ostrander, C., Dong, D., Shin, J., Presnell, S., Fox, B., Haldeman, B., Cooper, E., Taft, D., Gilbert, T., Grant, F.J., Tackett, M., Krivan, W., McKnight, G., Clegg, C., Foster, D., Klucher, K.M. (2003) IL-28, IL-29 and their class II cytokine receptor IL-28R. *Nat. Immunol.* 4:63-68.

253. Shinnick, T.M., Sutcliffe, J.G., Green, N., Lerner, R.A. (1983) Synthetic peptide immunogens as vaccines. *Ann. Rev. Microbiol.* 37:425-446.

254. Shorter, J., Beard, M.B., Seemann, J., Dirac-Svejstrup, A.B., Warren, G. (2002) Sequential tethering of Golgins and catalysis of SNAREpin assembly by the vesicle-tethering protein p115. *J. Cell Biol.* 157:45-62.

255. Siebenlist, U., Simpson, R.B., Gilbert, W. (1980) *E. coli* RNA polymerase interacts homologously with two different promoters. *Cell* 20:269-281.

256. Siegal, G.J., Agranoff, B.W., Albers, R.W., Fisher, S.K., Uhler, M.D., eds. (1999) Basic Neurochemistry, Molecular, Cellular, and Medical Aspects. 6th ed. Lippencott, Williams & Wilkins.

257. Sladek, R., Bader, J.A., Giguere, V. (1997) The orphan nuclear receptor estrogen-related receptor alpha is a transcriptional regulator of the human medium-chain acyl coenzyme A dehydrogenase gene. *Mol. Cell Biol.* 17:5400-5409.

258. Slavin, S., Or, R., Aker, M., Shapira, M.Y., Panigrahi, S., Symeonidis, A., Cividalli, G., Nagler, A. (2001) Nonmyeloablative stem cell transplantation for the treatment of cancer and life-threatening nonmalignant

disorders: past accomplishments and future goals. *Cancer Chemother. Pharmacol.* 48:S79-S84.

259. Smit, A.F., Riggs, A.D. (1996) Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci.* 93:1443-1448.

260. Smith, G.E., Ju, G., Ericson, B.L., Moschera, J., Lahm, H.W., Chizzonite, R., Summers, M.D. (1985) Modification and secretion of human interleukin 2 produced in insect cells by a baculovirus expression vector. *Proc. Natl. Acad. Sci.* 82:8404-8408.

261. Smith, T.F., Waterman, M.S. (1981) Comparison of biosequences. *Adv. Appl. Math.* 2:482-489.

262. Soejima, H., Kawamoto, S., Akai, J., Miyoshi, O., Arai, Y., Morohka, T., Matsuo, S., Niikawa, N., Kimura, A., Okubo, K., Mukai, T. (2001) Isolation of novel heart-specific genes using the BodyMap database. *Genomics*. 74:115-120.

263. Soulier, S., Vilotte, J.L., L'Huillier, P.J., Mercier, J.C. (1996) Developmental regulation of murine integrin beta 1 subunit- and Hsc73-encoding genes in mammary gland: sequence of a new mouse Hsc73 cDNA. *Gene* 172:285-289.

264. Southern, E., Mir, K., Shchepinov, M. (1999) Molecular interactions on microarrays. *Nature Genetics* 21:5-9.

265. Stein, C.A., Kreig, A.M., eds. (1998) Applied Antisense Oligonucleotide Technology. Wiley-Liss.

266. Steinhaur, C., Wingren, C., Hager, A.C., Borrebaeck, C.A. (2002) Single framework recombinant antibody fragments designed for protein chip applications. *Biotechniques, Supp.*:38-45.

267. Stetler-Stevenson, W.G., Liotta, L.A., Kleiner, D.E. Jr. (1993) Extracellular matrix 6: role of matrix metalloproteinases in tumor invasion and metastasis. *FASEB J.* 7:1434-1441.

268. Stewart, Z.A., Westfall, M.D., Pietenpol, J.A. (2003) Cell-cycle dysregulation and anticancer therapy. *Trends Pharmacol. Sci.* 24:139-145.

269. Sturm, A., Dignass, A.U. (2002) Modulation of gastrointestinal wound repair and inflammation by phospholipids. *Biochim. Biophys. Acta* 1582:282-288.

270. Stutz, F., Bachi, A., Doerks, T., Braun, I.C., Seraphin, B., Wilm, M., Bork, P., Izaurralde, E. (2000) REF, an evolutionary conserved family of hnRNP-like proteins, interacts with TAP/Mex67p and participates in mRNA nuclear export. *RNA* 6:638-650.

271. Suh, Y.H., Checler, F. (2002) Amyloid precursor protein, presenilins, and alpha-synuclein: molecular pathogenesis and pharmacological applications in Alzheimer's disease. *Pharmacol. Rev.* 54:469-525.

272. Sutcliffe, J.G., Shinnick, T.M., Green, N., Lerner, R.A. (1983) Antibodies that react with predetermined sites on proteins. *Science* 219:660-666.

273. Tan, J., Town, T., Paris, D., Mori, T., Suo, Z., Crawford, F., Mattson, M.P., Flavell, R.A., Mullan, M. (1999) Microglial activation resulting from CD40-CD40L interaction after beta-amyloid stimulation. *Science* 286:2352-2355.

274. Tekur, S., Pawlak, A., Guellaen, G., Hecht, N.B. (1999) Contrin, the human homologue of a germ-cell Y-box-binding protein: cloning, expression, and chromosomal localization. *J. Androl.* 20:135-144.

275. Terada, R., Yamamoto, K., Hakoda, T., Shimada, N., Okano, N., Baba, N., Ninomiya, Y., Gershwin, M.E., Shiratori, Y. (2003) Stromal cell-derived factor-1 from biliary epithelial cells recruits CXCR4-positive cells: implications for inflammatory liver diseases. *Lab. Invest.* 83:665-672.

276. Thompson, J.D., Higgins, D.G., Gibbon, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673-80.

277. Tilburn, J., Scazzocchio, C., Taylor, G.G., Zabicky-Zissman, J.H., Lockington, R.A., Davies, R.W. (1983) Transformation by integration in *Aspergillus nidulans*. *Gene* 26:205-221.

278. Trounson, A. (2002) Human embryonic stem cells: mother of all cell and tissue types. *Reprod. Biomed. Online* 4 Suppl. 1:58-63.

279. Tsuda, T., Gallup, M., Jany, B., Gum, J., Kim, Y., Basbaum, C. (1993) Characterization of a rat airway cDNA encoding a mucin-like protein. *Biochem. Biophys. Res. Commun.* 195:363-373.

280. Tukey, R.H., Pendurthi, U.R,, Nguyen, N.T., Green, M.D., Tephly, T.R. (1993) Cloning and characterization of rabbit liver UDP-glucuronosyltransferase cDNAs. Developmental and inducible expression of 4-hydroxybiphenyl UGT2B13. *J. Biol. Chem.* 268:15,260-15,266.

281. Vainberg, I.E., Lewis, S.A., Rommelaere, H., Ampe, C., Vandekerckhove, J., Klein, H.L., Cowan, N.J. (1998) Prefoldin, a chaperone that delivers unfolded proteins to cytosolic chaperonin. *Cell* 93:863-873.

282. Vale, R.D. (2003) The molecular motor toolbox for intracellular transport. *Cell* 112:467-480.

283. Vallejo, M., Ron, D., Miller, C.P., Habener, J.F. (1993) C/ATF, a member of the activating transcription factor family of DNA-binding proteins, dimerizes with CAAT/enhancer-binding proteins and directs their binding to cAMP response elements. *Proc. Natl. Acad. Sci.* 90:4679-4683.

284. van den Berg, J.A., van der Laken, K.J., van Ooyen, A.J., Renniers, T.C., Rietveld, K., Schaap, A., Brake, A.J., Bishop, R.J., Schultz, K., Moyer, D. (1990) *Kluyveromyces* as a host for heterologous gene expression: expression and secretion of prochymosin. *Bio/Technology* 8:135-139.

285. Van den Berghe, L., Laurell, H., Huez, I., Zanibellato, C., Prats, H., Bugler, B. (2000) FIF [fibroblast growth factor-2 (FGF-2)-interacting-factor], a nuclear putatively antiapoptotic factor, interacts specifically with FGF-2. *Mol. Endocrinol.* 14:1709-1724.

286. Van Den Blink, B., Ten Hove T., Van Den Brink G.R., Peppelenbosch M.P., Van Deventer S.J. (2002) From extracellular to intracellular targets, inhibiting MAP kinases in treatment of Crohn's disease. *Ann. N. Y. Acad. Sci.* 973:349-58.

287. van der Spoel, A.C., Jeyakumar, M., Butters, T.D., Charlton, H.M., Moore, H.D., Dwek, R.A., Platt, F.M. (2002) Reversible infertility in male mice after oral administration of alkylated imino sugars: a nonhormonal approach to male contraception. Proc. Natl. Acad. Sci. 99:17173-17178.

288. Van Eerdewegh, P., Little, R.D., Dupuis, J., Del Mastro, R.G., Falls, K., Simon, J., Torrey, D., Pandit, S., McKenny, J., Braunschweiger, K., Walsh, A., Liu, Z., Hayward, B., Folz, C., Manning, S.P., Bawa, A., Saracino, L., Thackston, M., Benchekroun, Y., Capparell, N., Wang, M., Adair, R., Feng, Y., Dubois, J., FitzGerald, M.G., Huang, H., Gibson, R., Allen, K.M.,

Pedan, A., Danzig, M.R., Umland, S.P., Egan, R.W., Cuss, F.M., Rorke, S., Clough, J.B., Holloway, J.W., Holgate, S.T., Keith, T.P. (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature*. 418:426-430.

289. Van Laar, J.M., Tyndall, A. (2003) Intense immunosuppression and stem-cell transplantation for patients with severe rheumatic autoimmune disease: a review. *Cancer Control* 10:57-65.

290. Verhey, K.J., Meyer, D., Deehan, R., Blenis, J., Schnapp, B.J., Rapoport, T.A., Margolis, B. (2001) Cargo of kinesin identified as JIP scaffolding proteins and associated signaling molecules. *J. Cell Biol.* 152:959-970.

291. Vlak, J.M., Klinkenberg, F.A., Zaal, K.J., Usmany, M., Klinge -Roode, E.C., Geervliet, J.B., Roosien, J.,van Lent, J.W. (1988) Functional studies on the p10 gene of Autographa californica nuclear polyhedrosis virus using a recombinant expressing a p10-beta- galactosidase fusion gene. *J. Gen. Virol.* 69:765-776.

292. Voisset, C., Bouton, O., Bedin, F., Duret, L., Mandrand, B., Mallet, F., Paranhos-Baccala. G. (2000) Chromosomal distribution and coding capacity of the human endogenous retrovirus HERV-W family. *AIDS Res. Hum. Retroviruses* 16:731-740.

293. Walker, J.E., Arizmendi, J.M., Dupuis, A., Fearnley, I.M., Finel, M., Medd, S.M., Pilkington, S.J., Runswick, M.J., Skehel, J.M. (1992) Sequences of 20 subunits of NADH:ubiquinone oxidoreductase from bovine heart mitochondria. Application of a novel strategy for sequencing proteins using the polymerase chain reaction. *J. Mol. Biol.* 226:1051-1072.

294. Walsh, A.C., Feulner, J.A., Reilly, A. (2001) Evidence for functionally significant polymorphism of human glutamate cysteine ligase catalytic subunit: association with glutathione levels and drug resistance in the National Cancer Institute tumor cell line panel. *Toxicol. Sci.* 61:218-223.

295. Wang, J., Kirby, C.E., Herbst, R. (2002) The tyrosine phosphatase PRL-1 localizes to the endoplasmic reticulum and the mitotic spindle and is required for normal mitosis. *J. Biol. Chem.* 277:46659-46668.

296. Wang, M.S., Schinzel, A., Kotzot, D., Balmer, D., Casey, R., Chodirker, B.N., Gyftodimou, J., Petersen, M.B., Lopez-Rangel, E., Robinson,

W.P. (1999) Molecular and clinical correlation study of Williams-Beuren syndrome: No evidence of molecular factors in the deletion region or imprinting affecting clinical outcome. *Am. J. Med. Genet.* 86:34-43.

297. Wax, S.D., Rosenfield, C.L., Taubman, M.B. (1994) Identification of a novel growth factor-responsive gene in vascular smooth muscle cells. *J. Biol. Chem.* 269:13,041-13,047.

298. Wei, S., Charmley, P., Concannon, P. (1997) Organization, polymorphism, and expression of the human T-cell receptor AV1 subfamily. *Immunogenetics* 45:405-412.

299. Weishaar, R.E., Cain, M.H., Bristol, J.A. (1985) A new generation of phosphodiesterase inhibitors: multiple molecular forms of phosphodiesterase and the potential for drug selectivity. *J. Med. Chem.* 28:537-545.

300. Weiner, H.L., Selkoe, D.J. (2002) Inflammation and therapeutic vaccination in CNS diseases. *Nature* 420:879-884.

301. Weinstein, M.E., Grossman, A., Perle, M.A., Wilmot, P.L., Verma, R.S., Silver, R.T., Arlin, Z., Allen, S.L., Amorosi, E., Waintraub, S.E., et al. (1988) The karyotype of Philadelphia chromosome-negative, bcr rearrangement-positive chronic myeloid leukemia. *Cancer Genet Cytogenet.* 35:223-229.

302. Weissman, I.L. (2000) Translating stem and progenitor cell biology to the clinic: barriers and opportunities. *Science* 287:1442-1446.

303. Weng, S., Gu, K., Hammond, P.W., Lohse, P., Rise, C., Wagner, R.W., Wright, M.C., Kuimelis, R.G. (2002) Generating addressable protein microarrays with PROfusion covalent mRNA-protein fusion technology. *Proteomics* 2:48-57.

304. Wenger, R.H., Rochelle, J.M., Seldin, M.F., Kohler, G., Nielsen, P.J. (1993) The heat stable antigen (mouse CD24) gene is differentially regulated but has a housekeeping promoter. *J. Biol. Chem.* 268:23,345-23,352.

305. Werner, T., Brack-Werner, R., Leib-Mosch, C., Backhaus, H., Erfle, V., Hehlmann, R. (1990) S71 is a phylogenetically distinct endogenous retroviral element with structural and sequence homology to mimian sarcoma virus (SSV). *Virology* 174:225-238.

306. Wick, G., Kromer, G., Neu, N., Fassler, R., Ziemiecki, A., Muller, R.G., Ginzel, M., Beladi, I., Kuhr, T., Hala, K. (1987) The multi-factorial pathogenesis of autoimmune disease. *Immunol. Lett.* 16:249-257.

307. Wieczorek, H., Brown, D., Grinstein, S., Ehrenfeld, J., Harvey, W.R. (1999) Animal plasma membrane energization by proton-motive V-ATPases. *Bioessays* 21:637-648.

308. Wieser, R. (2002) Rearrangements of chromosomal band 3q21 in myeloid leukemia. *Leuk. Lymphoma* 43:59-65.

309. Winssinger, N., Ficarro, S., Schultz, P.G., and Harris, J.L. (2002) Profiling protein function with small molecule microarrays. *Proc. Natl. Acad. Sci.* 99:11,139-11,144.

310. Wojtowicz-Praga, S. (1999) Clinical potential of matrix metalloprotease inhibitors. *Drugs R. D.* 1:117-129.

311. Wu, A.M., Gallo, R.C. (1975) Reverse Transcriptase. *CRC Crit. Rev. Biochem.* 3:289-347.

312. Yang, N., Shigeta, H., Shi, H., Teng, C.T. (1996) Estrogen-related receptor, hERR1, modulates estrogen receptor-mediated response of human lactoferrin gene promoter. *J. Biol. Chem.* 271:5795-5804.

313. Yelton, M.M., Hamer, J.E., Timberlake, W.E. (1984) Transformation of *Aspergillus* nidulans by using a trpC plasmid. *Proc. Natl. Acad. Sci.* 81:1470-1474.

314. Yoshihama, M., Uechi, T., Asakawa, S., Kawasaki, K., Kato, S., Higa, S., Maeda N., Minoshima, S., Tanaka, T., Shimizu, N., Kenmochi, N. (2002) The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome Res.* 12:379-390.

315. Yu, L., Zhang, Z., Loewenstein, P.M., Desai, K., Tang, Q., Mao, D., Symington, J.S., Green, M. (1995) Molecular cloning and characterization of a cellular protein that interacts with the human immunodeficiency virus type 1 Tat transactivator and encodes a strong transcriptional activation domain. *J. Virol.* 69:3007-3016.

316. Zallipsky, S. (1995) Functionalized poly(ethylene glycols) for preparation of biologically relevant conjugates. *Bioconjugate Chem.*, 6:150-165.

317. Zhang, Q., Acland, G.M., Wu, W.X., Johnson, J.L., Pearce-Kelling, S., Tulloch, B., Vervoort, R., Wright, A.F., Aguirre, G.D. (2002) Different RPGR exon ORF15 mutations in Canids provide insights into photoreceptor cell degeneration. *Hum. Mol. Genet.* 11:993-1003.

318. Zhang, W.M., Popova, S.N., Bergman, C., Velling, T., Gullberg, M.K., Gullberg, D. (2002) Analysis of the human integrin alpha11 gene (ITGA11) and its promoter. *Matrix Biol.* 21:513-523.

319. Zhao, H., Grabowski, G.A. (2002) Gaucher disease: Perspectives on a prototype lysosomal disease. *Cell Mol. Life Sci.* 59:694-707.

320. Zhao, N., Hashida, H., Takhshi, N., Misumi, Y., Sakaki, Y. (1995) High-density cDNA filter analysis: a novel approach for large-scale quantitative analysis of gene expression. *Gene* 156:207-215.

321. Zhu, H., Bilgin, M., Bangham, R., Hall,. D., Casamayor, P., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R.A., Gerstein, M., Snyder, M. (2001) Global analysis of protein activities using proteome chips. *Science* 293:2101-2105.

322. Zhu, H., Klemic, J.F., Chang, S., Bertone, P., Casamayor, A., Klemic, K.G., Smith, D., Gerstien, M., Reed, M.A., Snyder, M. (2000) Analysis of yeast protein kinases using protein chips. *Nat. Genetics* 26:283-289.

323. Zhu, H., Snyder, M. (2003) Protein chip technology. *Curr. Opin. Chem. Biol.* 7:55-63.

SEQUENCE LISTING

[0448] A sequence listing in electronic format accompanies this application.

CLAIMS

1. A first nucleic acid molecule comprising a polynucleotide sequence chosen from at least one polynucleotide sequence according to SEQ ID NOS.: 1-209; SEQ ID NOS.: 419-627, or a complement thereof, or from at least one polynucleotide sequence that encodes SEQ ID NOS: 210-418.

2. The nucleic acid molecule of claim 1, wherein the nucleic acid molecule is a DNA or a RNA molecule.

3. An animal injected with the nucleic acid molecule of claim 1.

4. A double-stranded isolated nucleic acid molecule comprising the first nucleic acid molecule of claim 1 and its complement.

5. The nucleic acid molecule of claim 4, wherein the first polynucleotide sequence encodes a polypeptide chosen from a polypeptide comprising a signal peptide, a mature polypeptide that lacks a signal peptide, a signal peptide, a biologically active fragment of a polypeptide, a polypeptide lacking a signal peptide cleavage site, a polypeptide consisting essentially of a N-terminal fragment that contains a Pfam domain, and a polypeptide consisting essentially of a C-terminal fragment that contains a Pfam domain.

6. A second nucleic acid molecule comprising a second polynucleotide sequence that is at least about 70%, or about 80%, or about 90%, or about 95% homologous to the first nucleic acid molecule of claim 1.

7. A second isolated nucleic acid molecule comprising a second polynucleotide sequence that hybridizes to the first polynucleotide sequence of claim 1 under high stringency conditions.

8. The second isolated nucleic acid molecule of claim 6, wherein the second polynucleotide sequence is complementary to the first polynucleotide sequence.

9. A vector comprising the nucleic acid molecule of claim 1 and a promoter that drives the expression of the nucleic acid molecule.

10. The vector of claim 9, wherein the promoter is chosen from one or more of a promoter that is naturally contiguous to the nucleic acid molecule, a promoter that is not naturally contiguous to the nucleic acid molecule, an inducible promoter, a conditionally active promoter, a constitutive promoter, and a tissue specific promoter.

11. A host cell transformed, transfected, transduced, or infected with the nucleic acid molecule of claim 1.

12. The host cell of claim 11, wherein the cell is chosen from one or more of a prokaryotic cell, a eucaryotic cell, a human cell, a mammalian cell, an insect cell, a fish cell, a plant cell, and a fungal cell.

13. A nucleic acid composition comprising a pharmaceutically acceptable carrier or a buffer and one or more compositions chosen from the nucleic acid molecule of claim 1, the nucleic acid molecule of claim 4, the vector of claim 9, and the host cell of claim 11.

14. One or more polypeptide molecules comprising a polypeptide sequence chosen from at least one amino acid sequence according to SEQ ID NOS.: 210-418.

15. An animal injected with the polypeptide molecule of claim 14.

16. The polypeptide of claim 14, wherein the polypeptide has a function chosen from an agonist, an antagonist, a ligand, and a receptor.

17. The polypeptide of claim 14, wherein the polypeptide is chosen from a polypeptide comprising a signal peptide, a mature polypeptide that lacks a signal peptide, a signal peptide, a biologically active fragment of a polypeptide, a polypeptide lacking a signal peptide cleavage site, a biologically active fragment consisting essentially of an N-terminal fragment containing a Pfam domain, and a C-terminal fragment containing a Pfam domain.

18. A polypeptide composition comprising the polypeptide molecule of claim 14 and a pharmaceutically acceptable carrier or a buffer.

19. A cell culture medium comprising the polypeptide of claim 14.

20. The cell culture medium of claim 19, further comprising responder cells chosen from one or more T cells, B cells, NK cells, dendritic cells, macrophages, muscle cells, stem cells, epithelial skin cells, fat cells, blood cells, brain cells, bone marrow cells, endothelial cells, retinal cells, bone cells, kidney cells, pancreatic cells, liver cells, spleen cells, prostate cells, cervical cells, ovarian cells, breast cells, lung cells, liver cells, soft tissue cells, colorectal cells, cells of the gastrointestinal tract, and cancer cells.

21. The cell culture medium of claim 20, wherein the responder cells proliferate in the medium.

22. The cell culture medium of claim 20, wherein the responder cells are inhibited in the medium.

23. A cell culture comprising transfected cells, wherein the transfected cells are transfected with the polynucleotide of claim 1.

24. The cell culture of claim 23, further comprising responder cells chosen from one or more T cells, B cells, NK cells, dendritic cells, macrophages, muscle cells, stem cells, epithelial skin cells, fat cells, blood cells, brain cells, bone marrow cells, endothelial cells, retinal cells, bone cells, kidney cells, pancreatic cells, liver cells, spleen cells, prostate cells, cervical cells, ovarian cells, breast cells, lung cells, liver cells, soft tissue cells, colorectal cells, cells of the gastrointestinal tract, and cancer cells.

25. The cell culture of claim 23, wherein the responder cells proliferate in the cell culture.

26. The cell culture of claim 23, wherein the responder cells are inhibited in the cell culture.

27. A method of making a transformed, transfected, transduced, or infected host cell comprising:

(a) providing a composition comprising the vector of claim 9, and

(b) allowing a host cell to come into contact with the vector to form a transformed, transfected, transduced, or infected host cell.

28. A method of making a polypeptide comprising:

(a) providing a nucleic acid molecule that comprises a polynucleotide sequence encoding the polypeptide of claim 14;

(b) introducing the nucleic acid molecule into an expression system; and

(c) allowing the polypeptide to be produced.

29. A method of making a polypeptide comprising:

(a) providing a composition comprising the host cell of claim 11;

(b) culturing the host cell to produce the polypeptide; and

(c) allowing the polypeptide to be produced.

30. A diagnostic kit comprising a polynucleotide molecule, wherein the polynucleotide molecule comprises a sequence chosen from (a) at least 6, (b) at least 7, (c) at least 8, and (d) at least 9 contiguous nucleotides chosen from the nucleic acid molecule of claim 1.

31. A diagnostic kit comprising a polypeptide molecule, wherein the polypeptide molecule comprises an amino acid sequence or a biologically active fragment thereof, derived from the nucleic acid molecule of claim 1.

32. A genetically modified mouse comprising a deletion, substitution, or modification of a sequence chosen from SEQ ID NOS.: 1-209; SEQ ID NOS.: 419-627, wherein the deletion, substitution or modification prevents or reduces expression of said sequence and results in a mouse deficient in or completely lacking one or more gene products of a sequence chosen from SEQ ID NOS.: 1-209; SEQ ID NOS.: 419-627.

33. A method of determining the presence of the nucleic acid molecule of claim 1 or its complement comprising:

(a) providing a complement to the nucleic acid molecule or providing a complement to the complement of the nucleic acid molecule;

(b) allowing the molecules to interact; and

(c) determining whether interaction has occurred.

34. A method of determining the presence of an antibody to the polypeptide of claim 14 in a sample, comprising:

(a) providing the polypeptide;

(b) allowing the polypeptide to interact with any specific antibody in the sample; and

(c) determining whether interaction has occurred.

35. A cell-free medium comprising the polypeptide of claim 14.

36. The cell-free medium of claim 35, further comprising lysates chosen from bacterial cells and eukaryotic cells.

37. The cell-free medium of claim 36, wherein the eukaryotic cells are wheat germ cells.

38. A non-human animal comprising the polynucleotide of claim 1, wherein the animal produces a human protein.

39. A non-human eukaryotic cell comprising the polynucleotide of claim 1, wherein the cell produces a human protein.

40. A bacterial cell comprising the polynucleotide of claim 1, wherein the cell produces a human protein.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 March 2004 (11.03.2004)

PCT

(10) International Publication Number
WO 2004/020595 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number:
PCT/US2003/027107

(22) International Filing Date: 28 August 2003 (28.08.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/406,616	29 August 2002 (29.08.2002)	US
60/406,655	29 August 2002 (29.08.2002)	US
60/406,640	29 August 2002 (29.08.2002)	US
60/406,576	29 August 2002 (29.08.2002)	US
60/406,666	29 August 2002 (29.08.2002)	US
60/406,611	29 August 2002 (29.08.2002)	US
60/406,612	29 August 2002 (29.08.2002)	US
60/411,019	17 September 2002 (17.09.2002)	US
60/411,024	17 September 2002 (17.09.2002)	US
60/411,046	17 September 2002 (17.09.2002)	US
60/411,082	17 September 2002 (17.09.2002)	US
60/411,022	17 September 2002 (17.09.2002)	US
60/410,962	17 September 2002 (17.09.2002)	US
60/410,953	17 September 2002 (17.09.2002)	US
60/410,957	17 September 2002 (17.09.2002)	US
60/411,037	17 September 2002 (17.09.2002)	US
60/410,951	17 September 2002 (17.09.2002)	US
60/410,946	17 September 2002 (17.09.2002)	US
60/410,960	17 September 2002 (17.09.2002)	US
60/411,111	17 September 2002 (17.09.2002)	US
60/411,052	17 September 2002 (17.09.2002)	US
60/463,732	18 April 2003 (18.04.2003)	US
60/463,700	18 April 2003 (18.04.2003)	US
60/467,230	2 May 2003 (02.05.2003)	US
60/467,203	2 May 2003 (02.05.2003)	US
60,471,306	19 May 2003 (19.05.2003)	US
60/472,420	22 May 2003 (22.05.2003)	US
60/476,609	9 June 2003 (09.06.2003)	US
60/485,223	8 July 2003 (08.07.2003)	US
60/485,325	8 July 2003 (08.07.2003)	US
60/486,480	14 July 2003 (14.07.2003)	US
60/486,960	15 July 2003 (15.07.2003)	US
60/493,573	8 August 2003 (08.08.2003)	US
60/493,370	8 August 2003 (08.08.2003)	US

(71) Applicants *(for all designated States except US)*: **FIVE PRIME THERAPEUTICS, INC.** [US/US]; 951 Gateway Boulevard, South San Francisco, CA 94080 (US). **RIKEN** The Institute of Physical and Chemical Research [JP/JP]; 2-1, Hirosawa, Wako,Saitama, 351-0198 (JP). **KABUSHIKI KAISHA DNAFORM** [JP/JP]; 3-35, Mita 1-chome, Minato-ku, Tokyo 108-0073 (JP).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **WILLIAMS, Lewis, T.** [US/US]; 125 Chapel Drive, Mill Valley, CA 94941 (US). **CHU, Keting** [US/US]; 2017 Easton Drive, Burlingame, CA 94010 (US). **LEE, Ernestine** [US/US]; 226 Arlington Avenue, kensington, CA 94707 (US). **HESTIR, Kevin** [/]; 226 Arlington Avenue, Kensington, CA 94707 (US).

(74) Agent: **GARRETT, Arthur, S.**; Finnegan, Henderson, Farabow, Garrett & Dunner, L.L.P., 1300 I Street, NW, Washington, DC 20005-3315 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PG, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *with sequence listing part of description published separately in electronic form and available upon request from the International Bureau*

(48) Date of publication of this corrected version:
15 July 2004

(15) Information about Correction:
see PCT Gazette No. 29/2004 of 15 July 2004, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NOVEL HUMAN POLYPEPTIDES ENCODED BY POLYNUCLEOTIDES

(57) Abstract: The invention provides novel polynucleotides, related polypeptides, related nucleic acid and polypeptide compositions, and related modulators, such as antibodies and small molecule modulators. The compositions of the invention are useful in treating proliferative disorders, e.g., cancers, and inflammatory, immune, bacterial, and viral disorders.

WO 2004/020595 A2